

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Mar 26 08:28:10 1997; MasPar time 20.44 Seconds  
Tabular output not generated.  
Title: >US-08-409-122-2  
Description: (1-507) from US08409122.pep  
Sequence: 1 MALWRPSDNTVYLPSPVAR.....APSATTSSKPAKRVVRARK 507  
Scoring table: PAM 150  
Gap 11  
Searched: 82182 seqs, 25727515 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir48  
1:ann1 2:ann2 3:ann3 4:unann1 5:unann2 6:unann3 7:unann4  
8:unann5 9:unann6 10:unann7 11:unann8 12:unann9 13:unann14:unrev  
Statistics: Mean 47.890; Variance 101.152; scale 0.473  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.  
SUMMARIES

\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.  
\*\*\*\*\*

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Mar 26 08:28:10 1997; MasPar time 20.44 Seconds  
Tabular output not generated.  
Title: >US-08-409-122-2  
Description: (1-507) from US08409122.pep  
Sequence: 1 MALWRPSDNTVYLPSPVAR.....APSATTSSKPAKRVVRARK 507  
Scoring table: PAM 150  
Gap 11  
Searched: 82182 seqs, 25727515 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir48  
1:ann1 2:ann2 3:ann3 4:unann1 5:unann2 6:unann3 7:unann4  
8:unann5 9:unann6 10:unann7 11:unann8 12:unann9 13:unann14:unrev  
Statistics: Mean 47.890; Variance 101.152; scale 0.473  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.  
SUMMARIES

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Mar 26 08:28:10 1997; MasPar time 20.44 Seconds  
Tabular output not generated.  
Title: >US-08-409-122-2  
Description: (1-507) from US08409122.pep  
Sequence: 1 MALWRPSDNTVYLPSPVAR.....APSATTSSKPAKRVVRARK 507  
Scoring table: PAM 150  
Gap 11  
Searched: 82182 seqs, 25727515 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir48  
1:ann1 2:ann2 3:ann3 4:unann1 5:unann2 6:unann3 7:unann4  
8:unann5 9:unann6 10:unann7 11:unann8 12:unann9 13:unann14:unrev  
Statistics: Mean 47.890; Variance 101.152; scale 0.473  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.  
SUMMARIES

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Mar 26 08:28:10 1997; MasPar time 20.44 Seconds  
Tabular output not generated.  
Title: >US-08-409-122-2  
Description: (1-507) from US08409122.pep  
Sequence: 1 MALWRPSDNTVYLPSPVAR.....APSATTSSKPAKRVVRARK 507  
Scoring table: PAM 150  
Gap 11  
Searched: 82182 seqs, 25727515 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir48  
1:ann1 2:ann2 3:ann3 4:unann1 5:unann2 6:unann3 7:unann4  
8:unann5 9:unann6 10:unann7 11:unann8 12:unann9 13:unann14:unrev  
Statistics: Mean 47.890; Variance 101.152; scale 0.473  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.  
SUMMARIES

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Mar 26 08:28:10 1997; MasPar time 20.44 Seconds  
Tabular output not generated.  
Title: >US-08-409-122-2  
Description: (1-507) from US08409122.pep  
Sequence: 1 MALWRPSDNTVYLPSPVAR.....APSATTSSKPAKRVVRARK 507  
Scoring table: PAM 150  
Gap 11  
Searched: 82182 seqs, 25727515 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir48  
1:ann1 2:ann2 3:ann3 4:unann1 5:unann2 6:unann3 7:unann4  
8:unann5 9:unann6 10:unann7 11:unann8 12:unann9 13:unann14:unrev  
Statistics: Mean 47.890; Variance 101.152; scale 0.473  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.  
SUMMARIES

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Mar 26 08:28:10 1997; MasPar time 20.44 Seconds  
Tabular output not generated.  
Title: >US-08-409-122-2  
Description: (1-507) from US08409122.pep  
Sequence: 1 MALWRPSDNTVYLPSPVAR.....APSATTSSKPAKRVVRARK 507  
Scoring table: PAM 150  
Gap 11  
Searched: 82182 seqs, 25727515 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir48  
1:ann1 2:ann2 3:ann3 4:unann1 5:unann2 6:unann3 7:unann4  
8:unann5 9:unann6 10:unann7 11:unann8 12:unann9 13:unann14:unrev  
Statistics: Mean 47.890; Variance 101.152; scale 0.473  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.  
SUMMARIES

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Mar 26 08:28:10 1997; MasPar time 20.44 Seconds  
Tabular output not generated.  
Title: >US-08-409-122-2  
Description: (1-507) from US08409122.pep  
Sequence: 1 MALWRPSDNTVYLPSPVAR.....APSATTSSKPAKRVVRARK 507  
Scoring table: PAM 150  
Gap 11  
Searched: 82182 seqs, 25727515 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir48  
1:ann1 2:ann2 3:ann3 4:unann1 5:unann2 6:unann3 7:unann4  
8:unann5 9:unann6 10:unann7 11:unann8 12:unann9 13:unann14:unrev  
Statistics: Mean 47.890; Variance 101.152; scale 0.473  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.  
SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	3704	99.5	568	3	P1WL18	L1 protein - human p 0.00e+00
2	3398	91.3	539	6	S36566	late protein - human 0.00e+00
3	3052	82.0	505	3	P1WL39	L1 protein - human p 0.00e+00
4	3030	81.4	503	3	P1WL39	L1 protein - human p 0.00e+00
5	2751	73.9	531	6	S36537	late protein - human 0.00e+00
6	2723	73.2	532	6	S36554	late protein - human 0.00e+00
7	2690	72.3	503	6	S36549	late protein - human 0.00e+00
8	2684	72.1	508	6	S36508	late protein - human 0.00e+00
9	2656	71.4	499	6	S36531	late protein - human 0.00e+00
10	2644	71.1	534	6	S36583	late protein - human 0.00e+00
11	2643	71.0	504	3	P1WL51	L1 protein - human p 0.00e+00
12	2624	70.5	504	3	P1WL31	L1 protein - human p 0.00e+00
13	2600	69.9	502	14	S36526	late protein - human 0.00e+00
14	2592	69.7	510	3	S15620	L1 protein - human p 0.00e+00
15	2589	69.6	531	3	P1WLHS	L1 protein - human p 0.00e+00
16	2567	69.0	510	3	S15627	L1 protein - human p 0.00e+00
17	2567	69.0	504	6	S36502	late protein - human 0.00e+00
18	2557	68.7	505	3	P1WL35	L1 protein - human p 0.00e+00
19	2549	68.5	524	3	P1WL58	L1 protein - human p 0.00e+00
20	2550	68.5	528	6	S36520	late protein - human 0.00e+00
21	2546	68.4	500	3	P1WL6	L1 protein - human p 0.00e+00

22	2547	68.4	505	6	S36589	late protein - human 0.00e+00
23	2540	68.3	499	3	P1WL33	L1 protein - human p 0.00e+00
24	2542	68.3	502	3	P1WL33	L1 protein - human p 0.00e+00
25	2542	68.3	505	6	S36560	late protein - human 0.00e+00
26	2526	67.9	529	6	S36578	late protein - human 0.00e+00
27	2520	67.7	499	3	P1WL13	L1 protein - human p 0.00e+00
28	2512	67.5	501	3	P1WL11	L1 protein - human p 0.00e+00
29	2484	66.8	503	6	S36514	late protein - human 0.00e+00
30	2475	66.5	502	3	P1WL42	L1 protein - human p 0.00e+00
31	1988	53.4	508	3	P1WL	L1 protein - human p 0.00e+00
32	1981	53.2	546	6	S36490	late protein - human 0.00e+00
33	1966	52.8	517	6	S36496	late protein - human 0.00e+00
34	1960	52.7	509	6	S36572	late protein - human 0.00e+00
35	1962	52.7	518	14	S36472	late protein - human 0.00e+00
36	1950	52.4	505	3	P1WL39	L1 protein - human 0.00e+00
37	1949	52.4	507	6	S36484	late protein - human 0.00e+00
38	1945	52.3	514	3	P1WL8	L1 protein - human p 0.00e+00
39	1938	52.1	517	6	S36543	late protein - human 0.00e+00
40	1932	51.9	507	6	S36478	late protein - human 0.00e+00
41	1923	51.7	507	6	S36595	late protein - human 0.00e+00
42	1917	51.5	514	3	P1WL47	L1 protein - human p 0.00e+00
43	1908	51.3	501	3	P1WL32	L1 protein - human p 0.00e+00
44	1906	51.2	501	3	P1WL32	L1 protein - human p 0.00e+00
45	1907	51.2	516	3	P1WL5	L1 protein - human p 0.00e+00

ALIGNMENTS

RESULT 1  
ENTRY #type complete  
TITLE L1 protein - human papillomavirus type 18  
ORGANISM #formal\_name human papillomavirus type 18  
DATE 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 24-Feb-1994  
ACCESSIONS A26251  
REFERENCE A92937  
#authors Cole, S.T.; Danos, O.  
#journal J. Mol. Biol. (1987) 193:599-608  
#title Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products.  
#cross-references MUID:87283882  
#accession A26251  
#molecule\_type DNA  
#residues 1-568 #label COL  
CLASSIFICATION #superfamily papillomavirus L1 protein  
KEYWORDS late protein  
SUMMARY #length 568 #molecular-weight 63741 #checksum 6794

```

#accession A26251
#cross references: EMBL:U7493862
#molecule_type DNA
##residues 1-568 ##label COL
CLASSIFICATION #superfamily papillomavirus L1 protein
KEYWORDS late protein
SUMMARY #length 568 #molecular-weight 63741 #checksum 6794

Query Match 99.5%; Score 3704; DB 3; Length 568;
Best Local Similarity 99.4%; Pred. No. 0.00e+00;
Matches 504; Conservative 1; Mismatches 2; Indels 0; Gaps 0

Db 62 malwrpsdntvylppspvarvntddvtptstifhagssrlltvgnpvfrvypaggnkq 121
QY 1 MALWRPSDNTVYLPSPVARVNTDDYTRTSIFHAGSSRLLTVCNPFYRVPAGGNKQ 60

Db 122 dipkvsaygyrfrvqlpdpnkfgldtsiynpctqrlvwacagveigrqplvglsgh 181
QY 61 DIPKVSAYGYRFRVQLPDPNFKFLPDNSIYNPQRLVWACAGVEIGRQPLVGLSGH 120

Db 182 pfynklidteshaatsnvsvdrnvsvdykqtdqclilgcapaigenwaktsrpl 241
QY 121 PFYNKLIDTESHAATSNVSEVDRNVSVDYKQTDQCLILGCAPAGEHWAKTSRPL 180

Db 242 sggdcpplelknviedgmvdtygandfstlqgtckevpldicskcykpylqmsad 301
QY 181 SGGDCPPLELKNVIEDGMVDTYGAMDFSLQDTKEVPLDICSCKYKPYLQMSAD 240

Db 302 pygdsmffcllrreqlfarhfnwragtmgdtvpqsllykgtgmraspgscvpspsgslv 361
QY 241 PYGDSMFFCLRREQLFARHFNWRAAGTMGDTVPQSLYKGTGMRASPGSCVPSPSGSIV 300

```

```

Db 362 tsdqlfknkpywlhkaqghnngcwhnqlfvtdtrstnlitcastgspvgqydatk 421
Qy 301 TSDQLFNKPYWLHKAQGHNGICWHNQLFVTVVDTTRSTNLITCASTGSPVPGQYDATK 360

Db 422 fkgysrhveeydlqfqlctitladvmayihmssilednwfvgvpppttslvdtr 481
Qy 361 FKQYSRHRVEYDLQFQLCTITLADVMAYIHSNMSSILEDNWFVGVPPTTSLVDTR 420

Db 482 fvgsvaltcokdaapeenkdydklfwnvdlkfkfslldoyplgrkflvqaglrkpt 541
Qy 421 FVGSVAITCOKDAAPENKDPYDKLFWNVDLKEKFSLLDQYPLGRKFLVQAGLRKPT 480

Db 542 lgpkrapsattsskpkrrvrzrk 568
Qy 481 lgpkrapsattsskpkrrvrzrk 507

RESULT 2
ENTRY #type complete
TITLE late protein - human papillomavirus type 45
ORGANISM #formal_name human papillomavirus type 45
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S36566
REFERENCE S36469
#authors Delius, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36566
#status preliminary
#molecule_type DNA
#residues 1-539 #label DEL
#cross-references EMBL:X74479
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 539 #molecular-weight 60310 #checksum 9613

Query Match 91.3%; Score 3398; DB 6; Length 539;
Best Local Similarity 87.3%; Pred. No. 0.00e+00;
Matches 448; Conservative 40; Mismatches 19; Indels 6; Gaps 4;

Db 27 malwrpsdntvylpppsavrvstddvyrtsifvfhagsrlltvgvnpvfyvngagnk 86
Qy 1 MALWRPSDNTVYLPSPSAVRVNTDDVYRTSIFVHAGSRLLTVGNPFRVPAGGNGK 59

Db 87 qavkvsayqyrvfrvalpdkfglpeditlynpetqrlwacvmeigrqplglglsq 146
Qy 60 QDPKVSAYQYRVFRVQLPDKFNKGLPNSIYNPETQRLWACAVEIGRQPLGLVLSG 119

Db 147 hpfynklodtesshaatavitqdvrdnsvdykqtlccllgcvpalgehkwagtklcpaq 206
Qy 120 HPFYNKLDOTESSHAATSNVEDVDNVDYKQTLCLLGCVPALGEHWAGTKLCPAQ 179

Db 207 lqpgdcpplelknltiledgmdvtdygamdfstlqdkcvepldicgickypdyiqmsa 266
Qy 180 LSGDCPPLELKNLTILEDGMDVTDYGMDFSTLQDKCEVPLDICQICKYPDYIQMSA 239

Db 267 dpygdmfcllrreqifarfhnrgvmgdtvptcdlyikgtseanrepvgcvyspspg 326
Qy 240 DPGDSMFCLLRREQIFARFHNWRAGTNGDTPVQSLYIKGTG--NRASPGSCVYSPPSG 297

Db 327 slitsdsqfknkpywlhkaqghnngicwhnqlfvtdtrstnlitcastqnpvpsdyd 386
Qy 298 SIVTSDSLFNKPYWLHKAQGHNGICWHNQLFVTVVDTTRSTNLITCASTQNPVPSDYD 357

Db 387 ptkfkqysrhveeydlqfqlctitlaevmsyihmssilenwfnvgvpppttslvd 446
Qy 358 ATKFKQYSRHRVEYDLQFQLCTITLADVMYSIHSNMSSILEDNWFVGVPPTTSLVD 417

Db 447 tyrfvqvavtcokdtkptekdpydklfkvtvdlkfkfslldoyplgrkflvqaglr 506
Qy 418 TYRFVQVAITCOKDAAPENKDPYDKLFKNVDLKEKFSLLDQYPLGRKFLVQAGLR 477

```

---

```

Db 507 rptgprkraeaststastaspkrvrzrk 539
Qy 478 KFTIGPKRRSAPSA-T--TSSKPKRRVRVRAR 507

RESULT 3
ENTRY #type complete
TITLE L1 protein - human papillomavirus type ME180 (provirus)
ORGANISM #formal_name human papillomavirus type ME180
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
27-Jan-1995
ACCESSIONS B40509
REFERENCE A40509
#authors H.; Schwarz, E.
#journal J. Virol. (1991) 65:5564-5568
#title Characterization of a novel human papillomavirus DNA in the cervical carcinoma cell line ME180.
#cross-references MUID:91374616
#accession B40509
#molecule_type DNA
#residues 1-505 #label REU
#cross-references GB:M73258
#note translation of the nucleotide sequence is not given
CLASSIFICATION #superfamily papillomavirus L1 protein
KEYWORDS late protein
SUMMARY #length 505 #molecular-weight 56805 #checksum 4496

Query Match 82.0%; Score 3052; DB 3; Length 505;
Best Local Similarity 76.7%; Pred. No. 0.00e+00;
Matches 386; Conservative 76; Mismatches 41; Indels 0; Gaps 0;

Db 1 malwrsdntvylpppsavrvstddvyrtrtgyvysgtrlltvgvnpvfyvngagrkq 60
Qy 1 MALWRPSDNTVYLPSPSAVRVNTDDVYTRTGYVYSGTRLLTVGNPFRVPAGGNGK 60

Db 61 dlpkvsayqyrvfrvalpdkfglpeditlynpetqrlwacvmeigrqplglglsq 120
Qy 61 DLPKVSAYQYRVFRVQLPDKFNKGLPNSIYNPETQRLWACAVEIGRQPLGLVLSG 120

Db 121 plynrlidatenspfesknpkardnvdykqtkclcgcvpalgehkwagkckpsnv 180
Qy 121 PFYNKLDOTESSHAATSNVEDVDNVDYKQTKCLCGCVPALGEHWAGTKCKSPPL 180

Db 181 qpgdcpplelntvpiqdgmditgygandfstlqetsevpldicgsvckypdyiqmsad 240
Qy 181 SQGDCPPLELNTVPIQDGMIDTGYGANDFSTLQETSEVPDLCQSVCKYPDYIQMSAD 240

Db 241 vvgdmfcllrreqifarfhnrgvmgdtlptelylkgtdirdspssyyvapspsgm 300
Qy 241 PYGDSMFCLLRREQIFARFHNWRAGTNGDTPVQSLYIKGTGMRASPGSCVYSPPSGSIV 300

Db 301 sdsqfknkpywlhkaqghnngicwhnqlfvtdtrstnlitsttesavpnlydpnk 360
Qy 301 TSDQLFNKPYWLHKAQGHNGICWHNQLFVTVVDTTRSTNLITCASTQSPVPGQYDATK 360

Db 361 fkeyrhrveeydlqfqlctitladvmayihmssilednwfvgvpppttslvdtr 420
Qy 361 FKQYSRHRVEYDLQFQLCTITLADVMAYIHSNMSSILEDNWFVGVPPTTSLVDTR 420

Db 421 ylgasaitcqkdapetkkpdyglnfwnvnlkfkfseidqfplgrkflvqaglrkpt 480
Qy 421 FVGSVAITCOKDAAPENKDPYDKLFKNVDLKEKFSLLDQYPLGRKFLVQAGLRKPT 480

Db 481 lgpkrapsattsskpkrrvrzrk 503
Qy 481 lgpkrapsattsskpkrrvrzrk 503

RESULT 4
ENTRY #type complete
TITLE L1 protein - human papillomavirus type 39

```

```

ORGANISM #formal_name human papillomavirus type 39
#note host Homo sapiens (man)
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
27-Jan-1995
ACCESSIONS H38502
REFERENCE A38502
#authors Volpers, C.; Streeck, R.E.
#journal Virology (1991) 181:419-423
#title Genome organization and nucleotide sequence of human papillomavirus type 39.
#cross-references MIMD:91135017
#accession H38502
#molecule_type DNA
#residues 1-505 #label VOL
#cross-references EMBL:M38185
#note translation of the nucleic acid sequence is not given
CLASSIFICATION #superfamily papillomavirus L1 protein
KEYWORDS late protein
SUMMARY #length 505 #molecular-weight 56604 #checksum 2488

Query Match 81.4%; Score 3030; DB 3; Length 505;
Best Local Similarity 76.7%; Pred. No. 0.00e+00;
Matches 389; Conservative 76; Mismatches 40; Indels 2; Gaps 2;

Db 1 mamwrsdmsvylpppsvakvntddvtrtgytgyagssrlltvgbpyfkvgmnggrkq 60
QY 1 MALWRPNDNTVYLPSPSVARVNTDDVTRTSIFYHAGSSRLLTGVNPFYRPVAGGNGKQ 60
Db 61 dipkvsayqrvfvtltpdnkfsipdaslynpetqrlrvacvvevgrgqplvglsgh 120
QY 61 DIPKVSAYQRVFVRLPDPNKFGLPNSINYPETQRLVWACAGVEIGRGOPLVGSLGH 120
Db 121 plynrqdgtens-pfsattnkdsrdnsydvkykqlcligcypaigehwkgkackpnnv 179
QY 121 PFYNKLDDESSHAATSNVSEDVNDVSDYKQQLCILGCAIPAIGEHNAKGTACKSRPL 180
Db 180 stgdcpplelvntpdegmdtgygamdfgalqetksevpldicqickypdyqlmsad 239
QY 181 SQGDCPPELEKNTVLEGDMDVGYGAMDFSTLQDTRCEVPLDICQICKYPPDYQLMSAD 240
Db 240 vygdsmfccrlreqlfarhfnwrgmvgdaipaqlylkgtdiranpssvycpssgm 299
QY 241 PYGDSMFFCLRRQLFAHFHFNWAGTMDVTPQSLYIKGTGMRASPGSCVYSPSPSGIV 300
Db 300 tsdsqfknkpywlhkgaghnngicwhnqlfvtvtdtrstnftstlesasipdygsk 359
QY 301 TSDSQLENKPYWLHKAQHNGNIGCHWNQLFVTVVDTTRSNLTICASTQSPVPCQYDATK 360
Db 360 fkeytrhveeydlqfqlctvtltdvmsyihthmssildnwnfavaappssaslvdytr 419
QY 361 FKQYSRHVEEYDLQFIFQLCTITLTADVMSYIHSNMSSILEDWNFGVPPPTTSLVDYR 420
Db 420 ylgasaitcqdapapekdpdygklkfwndlrkfealdqfqlgrkflqarvrirpt 479
QY 421 FVQSVAITCQKDAAPAKNKPDPYDKLFKNVNDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
Db 480 igprkpaastssss-atkhkrkrvsk 505
QY 481 IGPRKRSAPSATSSSKPAKRVRRARK 507

RESULT 5
ENTRY #type complete
TITLE late protein - human papillomavirus type 10
ORGANISM #formal_name human papillomavirus type 10
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S36537
REFERENCE S36469
#authors Delliuss, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36537

```

```

#status preliminary
#molecule_type DNA
#residues 1-531 #label DEL
#cross-references EMBL:X74465
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 531 #molecular-weight 59010 #checksum 1781

Query Match 73.9%; Score 2751; DB 6; Length 531;
Best Local Similarity 67.9%; Pred. No. 0.00e+00;
Matches 342; Conservative 103; Mismatches 57; Indels 2; Gaps 2;

Db 29 malwrsdnlvlypbtvpskvlstddvtrtnlyyyagstrlltvgbpyfipksnkv 88
QY 1 MALWRPNDNTVYLPSPSVARVNTDDVTRTSIFYHAGSSRLLTGVNPFYRPVAGGNGKQ 60
Db 89 dvpkvsayqrvfvtltpdnkfsipdaslynpetqrlrvacvvevgrgqplvglsgh 148
QY 61 DIPKVSAYQRVFVRLPDPNKFGLPNSINYPETQRLVWACAGVEIGRGOPLVGSLGH 120
Db 149 plynkldtenslnahgplgqdsrdnlsydvkykqlcligctppmgehkwgkgtcrrpp- 207
QY 121 PFYNKLDDESSHAATSNVSEDVNDVSDYKQQLCILGCAIPAIGEHNAKGTACKSRPL 180
Db 208 agdcpplelvntpdegmdtgygamdfgalqetksevpldicqickypdyqlmsae 267
QY 181 SQGDCPPELEKNTVLEGDMDVGYGAMDFSTLQDTRCEVPLDICQICKYPPDYQLMSAD 240
Db 268 vygdsmfccrlreqlfarhfnwrgmvgdaipdtfllksnggrdvgsavysptpsgm 327
QY 241 PYGDSMFFCLRRQLFAHFHFNWAGTMDVTPQSLYIKGTGMRASPGSCVYSPSPSGIV 300
Db 328 tseaqfknkpywlhkgaghnngicwhnqlfvtvtdtrstnftstlesasipdygsk 387
QY 301 TSDSQLENKPYWLHKAQHNGNIGCHWNQLFVTVVDTTRSNLTICASTQSPVPCQYDATK 360
Db 388 fkeytrhveeydlqfqlctvtltdvmsyihthmssildnwnfavaappssaslvdytr 447
QY 361 FKQYSRHVEEYDLQFIFQLCTITLTADVMSYIHSNMSSILEDWNFGVPPPTTSLVDYR 420
Db 448 flssasaitcqdapapekdpdygklkfwndlrkfealdqfqlgrkflqarvrirsa 507
QY 421 FVQSVAITCQKDAAPAKNKPDPYDKLFKNVNDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
Db 508 vsvrkripatsatgst-aakkrkrk 530
QY 481 IGPRKRSAPSATSSSKPAKRVRR 504

RESULT 6
ENTRY #type complete
TITLE late protein - human papillomavirus type 3
ORGANISM #formal_name human papillomavirus type 3
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S36554
REFERENCE S36469
#authors Delliuss, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36554
#status preliminary
#molecule_type DNA
#residues 1-532 #label DEL
#cross-references EMBL:X74462
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 532 #molecular-weight 59194 #checksum 1863

Query Match 73.2%; Score 2723; DB 6; Length 532;
Best Local Similarity 68.8%; Pred. No. 0.00e+00;
Matches 349; Conservative 91; Mismatches 60; Indels 7; Gaps 4;

Db 29 malwrsdnlvlypbtvpskvlstddvtrtnlyyyagstrlltvgbpyfipksnkv 88

```

```

Qy 1 MALWRPSDNTVYLPPLPPSVARVNTDDYVTRGSIIFYHAGSSRLLTGCVNPFVFPVAGGKNQ 60
Db 89 dlpkvsafgyvfvrlpdpnkfglpdparynpdaerlvwactgvevrglplvglsqh 148
Qy 61 DIPKVSAYQYRVRVQLPDPNKGFLPDNSIYNPETQRLVWACAGVEIGRGQPLGVLSGH 120
Db 149 plynklidatenslahdgkderdnisvdkqtkqlcvcgctppngehwkgtpckqna- 207
Qy 121 PFYNKLDDESSHAATSNVSEDYRDNVVDYKQQLCILCAPAIGEHAWGACKSRPL 180
Db 208 spgdcppelltapidqdmvdygandfnglsnksdvpidicctckypdyylmaae 267
Qy 181 SQGDCPPLELKNVTLVEDGDMDVDTGAGMDFSLQDTKCEVPLDICISCKTDPDLQMSAD 240
Db 268 pygdsmfylrkedlfarhflnragmagdtvpydalykgsdgsgrgklsavycptpsg 327
Qy 241 PYGDSMFCLRRREQLFARHFNWNRAGTMDGVTPQSLYIKG---TGMRASFGSCVYSPSPG 297
Db 328 smvtaetqlfnkpylrragghnngicwagqlfvvtvdttrstnmtlcvtstetst--yd 385
Qy 298 SIYVSDSOLFKNKPYLHKAQHNGNICHWNQLFVTVVDTTRSTNLTICASTQSPVPGQYD 357
Db 386 atkfkelylrhgeeydlqfqlckvcltpelmaylhtmnstlledwnfgltlpstaled 445
Qy 358 ATKFKQYSRHVEEYDLQFQLCTITLADVMYSIHSNNSILEDWNFGVPPPTTSLVD 417
Db 446 tyrfitsaictqkdapptekqpyaklnfvdvdlkdrfslqsfgrkflmqlgvgt 505
Qy 418 TYRFVQSVAITCQKDAAPAKNDPDKLKFNVDLKEKFSLDLDQYPLGRKFLVQAGLR 477
Db 506 rasiavkrsa-tttsrtaakrktk 531
Qy 478 KPTIGPKRSAPSATTSKPAKRVVR 504

RESULT 7
ENTRY S36549 #type complete
TITLE late protein human papillomavirus type 26
ORGANISM #formal_name human papillomavirus type 26
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995

ACCESSIONS S36549
REFERENCE S36469
#authors Delius, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36549
#status preliminary
#molecule_type DNA
#residues 1-503 #label DEL
#cross-references EMBL:X74472
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 503 #molecular-weight 56327 #checksum 9874

Query Match 72.3% Score 2690; DB 6; Length 503;
Best Local Similarity 67.3%; Pred. No. 0.00e+00;
Matches 337; Conservative 93; Mismatches 68; Indels 3; Gaps 3;

Db 1 malwrpsdntvylpdpnkfglpdparynpdaerlvwactgvevrglplvglsqh 59
Qy 1 MALWRPSDNTVYLPPLPPSVARVNTDDYVTRGSIIFYHAGSSRLLTGCVNPFVFPVAGGKNQ 60
Db 60 eipkvsayqyrvfvrhlpdpnkfglpdparynpdaerlvwactgvevrglplvglsqh 119
Qy 61 DIPKVSAYQYRVRVQLPDPNKGFLPDNSIYNPETQRLVWACAGVEIGRGQPLGVLSGH 120
Db 120 plfnklidatenslahatvnaetdnrvsvdnkqtkqlcigctppngehwgltckatqt 179
Qy 121 PFYNKLDDESSHAATSNVSEDYRDNVVDYKQQLCILCAPAIGEHAWGACKSRPL 180
Db 180 qrdcpelllssiedgmldgframdftalgetksdvpidisctckypdyylkmsad 239
Qy 181 SQGDCPPLELKNVTLVEDGDMDVDTGAGMDFSLQDTKCEVPLDICISCKTDPDLQMSAD 240

Query Match 72.1% Score 2684; DB 6; Length 508;
Best Local Similarity 67.1%; Pred. No. 0.00e+00;
Matches 337; Conservative 96; Mismatches 67; Indels 2; Gaps 2;

Db 8 mavrpsaetkylvlptpvsrvvtdyvttrtgiyyagssrlltlghpyfsipktg-gka 59
Qy 1 MALWRPSDNTVYLPPLPPSVARVNTDDYVTRGSIIFYHAGSSRLLTGCVNPFVFPVAGGKNQ 60
Db 68 dvpkvsafgyrvfvrhlpdpnkfglpdparynpdaerlvwactgvevrglplvgvsgn 127
Qy 61 DIPKVSAYQYRVRVQLPDPNKGFLPDNSIYNPETQRLVWACAGVEIGRGQPLGVLSGH 120
Db 128 plfnklidatenslanqdaesdrdnisvdpkqtkqlcigctppngehwgltckatqt 187
Qy 121 PFYNKLDDESSHAATSNVSEDYRDNVVDYKQQLCILCAPAIGEHAWGACKSRPL 180
Db 188 agqdcpplelknvtpidqdmvdygandfnglsnksdvpidisctckypdyylkmsad 247
Qy 181 SQGDCPPLELKNVTLVEDGDMDVDTGAGMDFSLQDTKCEVPLDICISCKTDPDLQMSAD 240
Db 248 aygdsmfylrreqlfarhfnragalgeclpstylykgttnrdppssvvyvatpsgm 307
Qy 241 PYGDSMFCLRRREQLFARHFNWNRAGTMDGVTPQSLYIKG---TGMRASFGSCVYSPSPG 300
Db 308 tseqlfnkpylrragghnngicwagqlfvvtvdttrstnmtlcvtstetst--yd 365
Qy 301 TSDSOLFKNKPYLHKAQHNGNICHWNQLFVTVVDTTRSTNLTICASTQSPVPGQYD 360
Db 366 ikqyrvhveeyelqfqlckvcltpelmaylhtmnstlledwnfgltlpstaled 425
Qy 361 FKQYSRHVEEYDLQFQLCTITLADVMYSIHSNNSILEDWNFGVPPPTTSLVDYTR 420

```



```

Db 426 yvksaitcqdqdpaeakelplakylkfdvnlqdsfslldqfplgrkflmqglvrtkps 485
Qy 421 FVQSVAITCQKDAAPAEKNDPYDLKFWNVLDKEKFSLDLQYPLGRKFLVQAGLRKPT 480
Db 486 tttkrsaspsststpsakrr 507
Qy 481 IGPKRRSAPSATTSSKPAKRRV 502

RESULT 9
ENTRY S36531 #type complete
TITLE late protein - human papillomavirus type 53
ORGANISM #formal_name human papillomavirus type 53
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S36531
REFERENCE S36469
#authors Dellus, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36531
#status preliminary
#molecule_type DNA
#residues 1-499 #label DEL
#cross-references EMBL:X7482
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 499 #molecular-weight 55722 #checksum 1339

Query Match 71.4%; Score 2656; DB 6; Length 499;
Best Local Similarity 66.3%; Pred. No. 0.00e+00;
Matches 333; Conservative 98; Mismatches 67; Indels 4; Gaps 3;

Db 1 mavrpsdskvlylptpvskvittdavvkrtrtiffhagssrllltvgpypysksg--ka 58
Qy 1 MALRPSDNTVYLPSPSVARVNTDDYTRISIEFHAGSSRLLVGNPFYFRVPPAGGNKQ 60
Db 59 dipkvsafqyrvfrirldpnpkfplpdtfnfndqerlhwacvglegrgqplgvsgsh 118
Qy 61 DIPKVSAYQYRVFRVQLPDPNPKFGLPDNSIYNPETQRLVWACAGVEIGRGQPLGVLSGH 120
Db 119 plfrldtesslaigtatpdsrdnsvdqkqtcqlciagcapaigehwtgktsrtp 178
Qy 121 PFYKRLDDTESSHAATSNVSDVRDVSVDYKQTLQILGCAPAI GEHWAAGTACKSRPL 180
Db 179 tagdcplleinspldgdmdvtgfainfkalgsksdvpldivqscckypdykmsad 238
Qy 181 SQGDCPPELEANTVLEDDGMDVDTGAGMDFSTLQDTKEVPLDICQSIKCKIPDYLQMSAD 240
Db 239 aygdsfmfyrrelqftrffragvigeelpndlylkgsgngrdpppsvvyvatpssgm 298
Qy 241 PYGDSMFCLRRELQFARHFNWAGTMDTVPSLYIKGTGMRASPGSCVTSPPSGSIV 300
Db 299 tseaqfknkpylgraghnngicwnqqlfvtvtdtrntamtisatqg--mst-ynskq 356
Qy 301 TSDSOLFKNKPYLHKAQHNNGICWHNQLFVTVVDTTSTNLITICASTQSPVPGQYDATK 360
Db 357 ikqvrvhaeeyleqfvlglckisaevmaylhtnmstllednwiglspvatsledkyr 416
Qy 361 FKQYSRHVEEDLQFIFQLCTITITADVMSYIHSNNSSILEDWNFGVPPPTTSLVDIYR 420
Db 417 yvksaaitcqdqdppekdplskylkfwvnlqnsfslldqfplgrkflmqgvrtkpp 476
Qy 421 FVQSVAITCQKDAAPAEKNDPYDLKFWNVLDKEKFSLDLQYPLGRKFLVQAGLRKPT 480
Db 477 vsskrsaspsststpsakrr 498
Qy 481 IGPKRRSAPSATTSSKPAKRRV 502

RESULT 10
ENTRY S36583 #type complete
TITLE late protein - human papillomavirus type 56
ORGANISM #formal_name human papillomavirus type 56

```

```

DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S36583
REFERENCE S36469
#authors Dellus, H.; Hofmann, B.
#submission submitted to the EMBL Data Library, August 1993
#description Primer-directed sequencing of human papillomavirus types.
#accession S36583
#status preliminary
#molecule_type DNA
#residues 1-534 #label DEL
#cross-references EMBL:X74483
CLASSIFICATION #superfamily papillomavirus L1 protein
SUMMARY #length 534 #molecular-weight 60160 #checksum 1336

Query Match 71.1%; Score 2644; DB 6; Length 534;
Best Local Similarity 67.0%; Pred. No. 0.00e+00;
Matches 337; Conservative 91; Mismatches 69; Indels 6; Gaps 6;

Db 36 matrpsenkvyllpvtvskvvdsvykrtsifvfhagssrllavghpypysv--tkdntkt 94
Qy 1 MALRPSDNTVYLPSPSVARVNTDDYTRISIEFHAGSSRLLVGNPFYFRVPPAGGNKQ 60
Db 95 nipkvsayqyrvfrirldpnpkfplpdtfnfndqerlhwacvglegrgqplgaglsgh 154
Qy 61 DIPKVSAYQYRVFRVQLPDPNPKFGLPDNSIYNPETQRLVWACAGVEIGRGQPLGVLSGH 120
Db 155 plfrldtesslannnviedsrdnsvdqkqtcqlciagcapaigehwtgkavckstqv 214
Qy 121 PFYKRLDDTESSHAATSNVSDVRDVSVDYKQTLQILGCAPAI GEHWAAGTACKSRPL 180
Db 215 ttgdcplleinspldgdmditgfamdkvlgskaeavpldivqscckypdykmsad 274
Qy 181 SQGDCPPELEANTVLEDDGMDVDTGAGMDFSTLQDTKEVPLDICQSIKCKIPDYLQMSAD 240
Db 275 aygdsfmfyrrelqftrffragvigeelpndlylkgsgngrdpppsvvyvatpssgm 334
Qy 241 PYGDSMFCLRRELQFARHFNWAGTMDTVPSLYIKGTGMRASPGSCVTSPPSGSIV 300
Db 335 tseaqfknkpylgraghnngicwnqqlfvtvtdtrntamtisatqg--qls-kydark 392
Qy 301 TSDSOLFKNKPYLHKAQHNNGICWHNQLFVTVVDTTSTNLITICASTQSPVPGQYDATK 360
Db 393 inqvlrvhaeeyleqfvlglckitlsaevmaylhtnmstllednwiglspvatsledkyr 452
Qy 361 FKQYSRHVEEDLQFIFQLCTITITADVMSYIHSNNSSILEDWNFGVPPPTTSLVDIYR 420
Db 453 yvrstaitcqdqdppekdplakylkfdvnlqdsfslldqfplgrkflmqglvrtkps 512
Qy 421 FVQSVAITCQKDAAPAEKNDPYDLKFWNVLDKEKFSLDLQYPLGRKFLVQAGLRKPT 480
Db 513 vatskksaspsstst--st-pakrr 533
Qy 481 IGPKRRSAPSATTSSKPAKRRV 502

RESULT 11
ENTRY P1MWJ51 #type complete
TITLE L1 protein - human papillomavirus type 51
ORGANISM #formal_name human papillomavirus type 51
#note host Homo sapiens (man)
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
27-Jan-1995
ACCESSIONS G40415
REFERENCE A40415
#authors Lungu, O.; Crum, C.P.; Silverstein, S.J.
#journal J. Virol. (1991) 65:4216-4225.
#title Biologic properties and nucleotide sequence analysis of human papillomavirus type 51.
#cross-references MUID:91303675
#accession G40415
#molecule_type DNA
#residues 1-504 #label LUN

```

##cross-references GB:M62877  
##note translation of the nucleotide sequence is not given  
CLASSIFICATION superfamily papillomavirus L1 protein  
KEYWORDS late protein  
SUMMARY #length 504 #molecular-weight 56314 #checksum 7345

Query Match 71.0%; Score 2643; DB 3; Length 504;  
Best Local Similarity 64.0%; Pred. No. 0.00e+00;  
Matches 323; Conservative 112; Mismatches 67; Indels 3; Gaps 3;

Db 1 malwrtndskvylppapvsvivnteevtrtgyayagssrltlghpyfpipkts-tra 59  
QY 1 MALWRPSDNTVYLPSPSVARVNTDDYVTRTSIFYHAGSSRLTLVGNPFRVP-AGGNGK 60  
Db 60 alpkaafgyrvfrvqlpdpnkfgldpdpnlypndtdrlvsgvgevsgqlvgvlgsh 119  
QY 61 DIPKVSAYQYRVRVQLPDPNKFGLPDNSIYNPETQRLVWACAGVEIGRQPLGVGLSGH 120  
Db 120 plfnkyddtensariangnagdvrdntsdvntkgtcltcligcappigehwgttckntpv 179  
QY 121 PFYNKLDDESSHAATSNVSEDVRDVSVDYKQTLCLGCAFAIGEHWAKTACKSRPL 180  
Db 180 ppqdcpllelvasvlgdgmldtfgamfaalqatksdvpldisgvckypdyikmsad 239  
QY 181 SOGDCCPPELKNVLEDDGMVDYTGAMDFSTLQDKCEVPLDQCISICKYFDYLOMSAD 240  
Db 240 tygnsmfflrreglfarhyvnlkvgedpndvyikgsgngrdpliesviyatspsgsm 299  
QY 241 PYGDSMFFCLRREQLFARHFNWAGTMDTVPSQSLYIKGTGM-RASPGSCVYSPSPSGSI 299  
Db 300 itadsgifnkpywlrhagghngicwngnqlfitevdttrstnlticstaaavspftfps 358  
QY 300 VTSDSOLFKNPYWLHKAQGHNGICWNLQFLVTVVDTTRSTNLTCASQSPVPQGYDAT 359  
Db 359 nfkqylrhgeevlqfifqcltlttemavylhmdptileqwnfgltlpsaaleday 418  
QY 360 KFKQYSRHYEEDLQFIFQLCTITLADVMSYIHSNSSLIEDMNFVGPPTTSLVDY 419  
Db 419 rfnrnaatcqdtpqspkdpklakvfdvdkerfeldldgdfalgrkflilgvqrk 478  
QY 420 RFVQSVAITCQDAAPAKNDPYDKLKFNVNLDKFKFSLDLQYPLGRKFLVQGLRRKP 479  
Db 479 rpglkprassasasasakrrkvk 503  
QY 480 TIGPRKRSAPSATTSSKPAKRVVR 504

RESULT 12  
ENTRY #type complete  
TITLE L1 protein - human papillomavirus type 31  
ORGANISM #formal\_name human papillomavirus type 31  
#note host Homo sapiens (man)  
DATE 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 24-Feb-1994

ACCESSIONS G32444  
REFERENCE A94398  
#authors Goldsborough, M.D.; Disilvestre, D.; Temple, G.F.; Lorincz, A.T.  
#journal Virology (1989) 171:306-311  
#title Nucleotide sequence of human papillomavirus type 31: a cervical neoplasia-associated virus.

#cross-references MUID:89299478  
#accession G32444  
#molecule\_type DNA  
#residues 1-504 #label GOL  
COMMENT The DNA sequence was obtained from GenBank, release 61.0.  
CLASSIFICATION superfamily papillomavirus L1 protein  
KEYWORDS late protein  
SUMMARY #length 504 #molecular-weight 56352 #checksum 7380

Query Match 70.5%; Score 2624; DB 3; Length 504;  
Best Local Similarity 65.1%; Pred. No. 0.00e+00;  
Matches 329; Conservative 102; Mismatches 71; Indels 3; Gaps 3;

Db 1 mslwrseatevylppvpvskvsvstdevytrtnlyyhagsarllltvghpyvsiipksdnpkk 60  
QY 1 MALWRPSDNTVYLPSPSVARVNTDDYVTRTSIFYHAGSSRLTLVGNPFRVP-AGGNGK 59  
Db 61 lvpvksvqlgyrvfrvqlpdpnkfgldtksfynpbtqrlvwacvglevgqplvgvlgsh 120  
QY 60 QDIPKVSAYQYRVRVQLPDPNKFGLPDNSIYNPETQRLVWACAGVEIGRQPLGVGLSG 119  
Db 121 hollnkldtensariangnagdvrdntsdvntkgtcltcligcappigehwgttckntpv 180  
QY 120 HPFYNKLDDESSHAATSNVSEDVRDVSVDYKQTLCLGCAFAIGEHWAKTACKSRP 179  
Db 181 itpdcpllelvasvlgdgmldtfgamfaalqatksdvpldisgvckypdyikmsa 240  
QY 180 LSQGDCCPPELKNVLEDDGMVDYTGAMDFSTLQDKCEVPLDQCISICKYFDYLOMSA 239  
Db 241 epyydtlffylrrreglfarhyvnlkvgedpndvyikgsgngrdpliesviyatspsgsm 300  
QY 240 PFYGDMSMFFCLRREQLFARHFNWAGTMDTVPSQSLYIKGTGM-RASPGSCVYSPSPSGSI 299  
Db 301 vtsdaqifnkpywlrhagghngicwngnqlfitevdttrstnlticstaaavspftfps 359  
QY 300 VTSDSOLFKNPYWLHKAQGHNGICWNLQFLVTVVDTTRSTNLTCASQSPVPQGYDAT 359  
Db 360 nfkqylrhgeevlqfifqcltlttemavylhmdptileqwnfgltlpsaaleday 419  
QY 360 KFKQYSRHYEEDLQFIFQLCTITLADVMSYIHSNSSLIEDMNFVGPPTTSLVDY 419  
Db 420 rfvtsqaitcqdtpqspkdpklakvfdvdkerfeldldgdfalgrkflilgvqrk 479  
QY 420 RFVQSVAITCQDAAPAKNDPYDKLKFNVNLDKFKFSLDLQYPLGRKFLVQGLRRKP 479  
Db 480 kfkagkraspsattsskpaakrrkvk 503  
QY 480 TIGPRKRSAPSATTSSKPAKRVVR 504

RESULT 13  
ENTRY #type complete  
TITLE late protein - human papillomavirus type 35  
ORGANISM #formal\_name human papillomavirus type 35  
DATE 09-Dec-1993 #sequence\_revision 09-Dec-1993 #text\_change 09-Dec-1993

ACCESSIONS S36526  
REFERENCE S36469  
#authors Delius, H.; Hofmann, B.  
#submission submitted to the EMBL Data Library, August 1993  
#accession S36526  
#status preliminary  
#residues 1-502 #label DEL  
#cross-references EMBL:X7477

SUMMARY #length 502 #molecular-weight 56148 #checksum 2108

Query Match 59.9%; Score 2600; DB 14; Length 502;  
Best Local Similarity 65.1%; Pred. No. 0.00e+00;  
Matches 329; Conservative 95; Mismatches 78; Indels 3; Gaps 3;

Db 1 mslwrseatevylppvpvskvsvstdevytrtnlyyhagsarllltvghpyvsiipksdnpkk 60  
QY 1 MALWRPSDNTVYLPSPSVARVNTDDYVTRTSIFYHAGSSRLTLVGNPFRVPAGGNGK 60  
Db 61 avpkvsglgyrvfrvqlpdpnkfgldtksfynpbtqrlvwacvglevgqplvgvlgsh 120  
QY 61 DIPKVSAYQYRVRVQLPDPNKFGLPDNSIYNPETQRLVWACAGVEIGRQPLGVGLSGH 120  
Db 121 pllnkldtensariangnagdvrdntsdvntkgtcltcligcappigehwgttckntpv 180  
QY 121 PFYNKLDDESSHAATSNVSEDVRDVSVDYKQTLCLGCAFAIGEHWAKTACKSRPL 180  
Db 181 kagecppllelntvlgdgmldtfgamfaalqatksdvpldisgvckypdyikmsa 240  
QY 181 SOGDCCPPELKNVLEDDGMVDYTGAMDFSTLQDKCEVPLDQCISICKYFDYLOMSAD 240

```

Db 241 pygdmllfyllrreqmfvrlhfragvtgvetpadlyikgttg-tlpsts-yftpsgsmv 298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 241 PYGDSMFFCLRRRLQFLARHFNWNRAGTMDTVPGSLYIKGTGMRASPGSCVYSPSPSGSIV 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 299 tsdaqlfnkpylqragghnngicwensqlfvtvdttrtnmsvcsavass-dstyknkn 357
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 301 TSDSQLFNKPYPYLHKAQGNHNGICWNLQFLFVTVVDTTRTNLTICASTQSPVPGQYDATK 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 358 fkeylrhgeeydlqfllqckltladvmtvthsmnpslledwnfqltpppsgtledtyr 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 361 FKQYSRHVEYDLQFLQCTITLTADVMSYIHSNMSSILEDNWFGVPPPTSLVDITYR 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 418 ytsqavtcqkpsakpkdpknyfwevdlkekfsadldqfplgrkfilqaglkarpn 477
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 421 FVQSVAITCQKDAAPAKENKDPYDKLAFWNVDLKEKFSLDLDQYPLGRKFLVQAGLRKRT 480
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 478 frlgkraapastekskstkrkvks 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 481 IGPKRSAPSATTSSKPAKRVRA 505
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 14
ENTRY S15620 #type complete
TITLE L1 protein - human papillomavirus type 2a
ORGANISM #formal_name human papillomavirus type 2a
#note host Homo sapiens (man)
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
17-Feb-1994
ACCESSIONS S15620
REFERENCE S15614
#authors Hirsch-Behnam, A.; Delius, H.; de Villiers, E.M.
#journal Virus Res. (1990) 18:81-98
#title A comparative sequence analysis of two human papillomavirus
(HPV) types 2a and 57.
#cross-references MUID:91188699
#accession S15620
#molecule_type DNA
#residues 1-510 #label HIR
#cross-references EMBL:X55964
CLASSIFICATION #superfamily papillomavirus L1 protein
KEYWORDS late protein
SUMMARY #length 510 #molecular-weight 57193 #checksum 5538

Query Match 69.7%; Score 2592; DB 3; Length 510;
Best Local Similarity 64.1%; Pred. No. 0.00e+00;
Matches 323; Conservative 111; Mismatches 59; Indels 11; Gaps 9;

Db 17 malwrpnskvylpvtpskvistdvyrtnvvyhggssrlltvghpysikkkn-nkv 75
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1 MALWRPSDNTVYLPSPSVARVNTDYYRTSIFVHAGSSRLLTGVPYFRVPAGGNGK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 76 avpkvsgyrvfhwlpdpnkfslpadlydpdtrllwacvgevgrgqlgvvgvsh 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 61 DIPKVSAYQYRFRVQLPDPNFKGLPDNSIYNPQRLVWACAGVEIGRQPLGVLSGH 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 136 pynrldttenah--tpdtaddrenismdykqtqfllgckppigehwsktctngss- 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 121 PFYNKLDDESSHAATSNVSEDRNVSDYKQTQCLIGCAPAIGERHAWAKTACKSRPL 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 193 aagdcplqfnttiedgdmvetfgaldfatlqnsksdvpdictntckypdykmaae 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 181 SOGDCPPELKNVLEDDGMDVDTGYGAMDFSLQDKCEVPLDQCOSICKYDYQLQMSAD 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 253 pygdsdffllrreqmfrhfnlghgmgtltpdelyktsv-ptpgshvtytstpsgsmv 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 241 PYGDSMFFCLRRRLQFLARHFNWNRAGTMDTVPGSLYIKGTGMRASPGSCVYSPSPSGSIV 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 312 sseaglfinkpylrragghnngicwensqlfvtvdttrtnsvlca-teas-dtykatn 369
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 301 TSDSQLFNKPYPYLHKAQGNHNGICWNLQFLFVTVVDTTRTNLTICASTQSPVPGQYDATK 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 370 fkeylrhmeeydlqfllqckltltltpelmaylhmndpqlledwnfvgvpppsasldtyr 429
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Qy 361 FKQYSRHVEYDLQFLQCTITLTADVMSYIHSNMSSILEDNWFGVPPPTSLVDITYR 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 430 ylgasaitcqktpktpdtpdyasltfwdvdlisesfmsmdldqfplgrkfilrga--mpt 487
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 421 FVQSVAITCQKDAAPAKENKDPYDKLAFWNVDLKEKFSLDLDQYPLGRKFLVQAGLRKRT 480
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 488 vs-rkraavsgttptt-skrkivr 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 481 IGPKRSAPSATTSSKPAKRVRA 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 15
ENTRY PIWLHS #type complete
TITLE L1 protein - human papillomavirus type 16
ORGANISM #formal_name human papillomavirus type 16
DATE 28-May-1986 #sequence_revision 28-May-1986 #text_change
02-May-1994
ACCESSIONS A03640
REFERENCE A23355
#authors Seedorf, K.; Krammer, G.; Durst, M.; Suhai, S.; Roweckamp,
W.G.
#journal Virology (1985) 145:181-185
#title Human papillomavirus type 16 DNA sequence.
#cross-references MUID:85246220
#accession A03640
#molecule_type DNA
#residues 1-531 #label SEE
CLASSIFICATION #superfamily papillomavirus L1 protein
KEYWORDS late protein
SUMMARY #length 531 #molecular-weight 59554 #checksum 4969

Query Match 69.6%; Score 2589; DB 3; Length 531;
Best Local Similarity 65.1%; Pred. No. 0.00e+00;
Matches 327; Conservative 97; Mismatches 74; Indels 4; Gaps 4;

Db 27 malwlpseatvylppvpvskvstdeyvartniyhagtsrllavghpfpikkkpnnk1 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1 MALWRPSDNTVYLPSPSVARVNTDYYRTSIFVHAGSSRLLTGVPYFRVPAGGNGK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 87 lvpkvsqlyrvfwhlpdpnkfsgdtsfynpdktrllwacvgevgrgqlgvvgish 146
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 61 DIPKVSAYQYRFRVQLPDPNFKGLPDNSIYNPQRLVWACAGVEIGRQPLGVLSGH 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 147 pllnkldttenasayaanaagvndrecismdykqtqcllglckppigehwkgspctnvav 206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 121 PFYNKLDDESSHAATSNVSEDRNVSDYKQTQCLIGCAPAIGERHAWAKTACKSRPL 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 207 nsgdcppllelntviqgdgmvtgfgamdfllqanksevpdictsickypdykmaae 266
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 181 SOGDCPPELKNVLEDDGMDVDTGYGAMDFSLQDKCEVPLDQCOSICKYDYQLQMSAD 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 267 pygdsalfyllrreqmfvrlhfragvtgvenpddlyikgsgstanlassanyftpsgsmv 326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 241 PYGDSMFFCLRRRLQFLARHFNWNRAGTMDTVPGSLYIKGTGMRASPGSCVYSPSPSGSIV 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 327 tsdaqlfnkpylqragghnngicwensqlfvtvdttrtnmslcaaatsett-ykntn 385
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 301 TSDSQLFNKPYPYLHKAQGNHNGICWNLQFLFVTVVDTTRTNLTICASTQSPVPGQYDATK 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 386 fkeylrhgeeydlqfllqckltladvmtvthsmnstiledwnfqlpppggtledtyr 445
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 361 FKQYSRHVEYDLQFLQCTITLTADVMSYIHSNMSSILEDNWFGVPPPTSLVDITYR 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 446 fvtalacqkhtppapkeddpkkytfwevnlkekfsadldqfplgrkfilqaglkakp 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 421 FVQSVAITCQKDAAPAKENKDPYDKLAFWNVDLKEKFSLDLDQYPLGRKFLVQAGLRKRT 479
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 505 kftlgkrkatpstsststtakr 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 480 TIG-PRKRSAPSATTSSKPAKR 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Mon Mar 31 09:21:05 1997

US-08-409-122-2.rpr

Page 8

Search completed: Wed Mar 26 08:29:01 1997  
Job time : 51 secs.

\_\_\_\_\_

\*\*\*\*\*  
M P S R C H  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Mar 26 08:29:18 1997; Maspar time 11.98 Seconds  
Tabular output not generated. 435.828 Million cell updates/sec  
Title: >US-08-409-122-2  
Description: (1-507) from US08409122.pep  
Perfect Score: 3721  
Sequence: 1 MALWRPSDNTVYLPSPVAR.....APSATTSSKPAKRVVRARK 507

Scoring table: PAM 150  
Gap 11  
Searched: 88003 seqs, 10295656 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: a-geneseq25  
1:part1 2:part3 3:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
Statistics: Mean 35.688; Variance 147.981; scale 0.241

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2610	70.1	505	8	R38807	4.27e-252
2	2595	69.7	505	8	R38808	1.52e-250
3	2547	68.4	532	17	R30316	1.41e-245
4	1992	53.5	508	4	P30315	2.94e-188
5	1862	50.0	495	3	P51099	7.26e-175
6	1178	31.7	215	16	R88267	9.86e-105
7	764	20.5	221	16	R88273	8.54e-63
8	754	20.3	222	16	R88268	8.62e-62
9	742	19.9	220	16	R88269	1.38e-60
10	728	19.6	224	16	R88271	3.50e-59
11	724	19.5	224	16	R88274	8.80e-59
12	724	19.5	225	16	R88270	8.80e-59
13	717	19.3	220	16	R88272	4.42e-58
14	720	19.3	228	16	R88275	2.21e-58
15	466	12.5	88	6	R32057	3.33e-33
16	466	12.5	88	3	R14299	3.33e-33
17	292	7.8	39	1	P82074	1.66e-16
18	259	7.0	51	6	R32062	1.94e-13
19	259	7.0	51	6	R32059	1.94e-13
20	241	6.5	36	6	R32071	8.70e-12
21	226	6.1	36	6	R32065	2.01e-10

22	217	5.8	35	6	R32063	Human IgA reactive HP	1.30e-09
23	211	5.7	46	6	R32070	VLP reactive HPV16 L1	4.48e-09
24	193	5.2	35	6	R32066	VLP reactive HPV16 L1	1.77e-07
25	184	4.9	39	1	P82075	Hman papilloma virus	1.09e-06
26	164	4.4	26	6	R32068	VLP reactive HPV16 L1	5.79e-05
27	161	4.3	20	1	R04980	Papilloma virus type	1.04e-04
28	157	4.2	37	3	R14301	Seroreactive epitope	2.27e-04
29	151	4.1	20	1	R04966	Papilloma virus type	7.24e-04
30	146	3.9	36	6	R32069	VLP reactive HPV16 L1	1.89e-03
31	137	3.7	20	1	R04984	Papilloma virus type	1.04e-02
32	136	3.7	20	1	R04981	Papilloma virus type	1.25e-02
33	137	3.7	21	3	R15589	Immunopeptide derived	1.04e-02
34	134	3.6	20	1	R04965	Papilloma virus type	1.82e-02
35	133	3.6	88	3	R14298	Seroreactive epitope	2.19e-02
36	130	3.5	20	1	R04985	Papilloma virus type	3.82e-02
37	130	3.5	20	1	R04974	Papilloma virus type	3.82e-02
38	129	3.5	20	1	R04970	Papilloma virus type	4.59e-02
39	127	3.4	20	1	R04964	Papilloma virus type	6.54e-02
40	126	3.4	20	1	R04975	Papilloma virus type	7.97e-02
41	124	3.3	20	1	R04990	Papilloma virus type	1.15e-01
42	123	3.3	20	1	R04976	Papilloma virus type	1.38e-01
43	119	3.2	20	1	R04961	Papilloma virus type	2.84e-01
44	118	3.2	22	3	R14302	Seroreactive epitope	3.40e-01
45	115	3.1	20	6	R32058	Human IgG reactive HP	5.81e-01

ALIGNMENTS

RESULT 1  
ID R38807 standard; Protein; 505 AA.  
AC R38807;  
DT 21-JAN-1994 (first entry)  
DE BPV1 L1  
KW L1; capsid protein; bovine; human; papillomavirus; baculovirus;  
KW transfer vector; promoter; capsid protein; transformation; vaccine;  
KW neutralising antibody; vertebrate.  
OS Bovine papillomavirus.  
PN US032869-A.  
PD 15-JUL-1993.  
PF 03-SEP-1992; 941371.  
PR 03-SEP-1992; US-941371.  
PR 16-MAR-1993; US-032869.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
PI Kirnbauer R, Lowy DR, Schiller JT;  
DR WPI: 93-249995/31.  
DR N-PSDB; Q47166.  
PT Recombinant papilloma virus capsid proteins - for vaccines  
PT against papilloma virus and for diagnosis of virus infection  
PS Example 1; Page 32-34; 45pp; English.  
CC The sequences given in R38807-08 represent the L1 capsid proteins from  
CC bovine and human papillomavirus respectively. The DNA encoding these  
CC sequences may be inserted into a baculovirus transfer vector and  
CC operatively expressed by a promoter of the vector, and the capsid  
CC protein produced by transformed cells. These capsid proteins may be  
CC used in vaccines to induce high-titre neutralising antibody response  
CC in vertebrates.  
SQ Sequence 505 AA;  
  
Query Match 70.1%; Score 2610; DB 8; Length 505;  
Best Local Similarity 65.5%; Pred. No. 4.27e-252;  
Matches 328; Conservative 96; Mismatches 75; Indels 2; Gaps 2;  
  
Db 1 msilwpsatvylppvpskvstdevatnlyvhagtarllavghpyfplkpnanki 60  
QY 1 MALWRPSDNTVYLPSPVARVNTDDYTRTSIFYHAGSRLTLTVGNPYFRVPAAGGNKQ 60  
Db 1 lvpkvsqlyrvfihlpdnpkfgpdtfsyvdptqrlwacvvgvgrgqplvgvlgsh 120  
QY 1 DIPKVSAYQVRFVQLPDPNKGFLPDNSYNFETQRLVWACAGVEICRGQPLVGLSGH 120  
Db 1 pllnklldtenasayaagvndrecismdykqtqlclgckppighwgkspctnav 180  
QY 121 PFYNKLDTESSHAATSNVSEDVRDNDVYKQTKQLCILGCAPIGHEWAKGTACKSRPL 180



Qy 180 LSQGDCCPPLKNTVLEGGDMVDVGYGAMDFSTLQDTKEVPLDTCQICKYPPDYQLQMSA 239  
 Db 268 dpygdlrffirkegmfarhnfagevgevpdpdtliiksgnrtsvsgsiyvnptpsgl 327  
 Qy 240 DPGDSMFFCLRRQLFARHFNWNRAGTWGDIVPQSLYIKG-TGMRASPGSCVYSPSPSGI 299  
 Db 328 vsseaelfnkpywlqacghangicwgnqlfvtvvdtrstnmtlcaav-tt-astyns 385  
 Qy 300 VTSQSLFNKPYLWHLKAQHNGNHCWNLQFLVTVVDTTSTNLITICASTQSPVPGQYDAT 359  
 Db 386 dykymhrveeydqfifqlcsittisaemaylhtmpnsvledwfnlgspnngtledty 445  
 Qy 360 KFKQYSRHEVEYDQFIFQLCTITLTADVMSYIHSNMSSILEDWNFVPPPTTSLVDY 419  
 Db 446 ryvsgaitcqtpekekpdpkynslfwelnlkefseildqyplgrkfkllqsgyirs 505  
 Qy 420 RFVQSVAITCQDAAPENKDPYDKLKFVNDLKEKFSLDLDQYPLGRKFLVQAGLRKP 479  
 Db 506 sirtgvkr-pavskasaapkrakrkrk 532  
 Qy 480 TIGPKRSAPSATTSSKPAKRVVRARK 507

## RESULT 4

ID P30315 standard; Protein; 508 AA.  
 AC P30315;  
 DT 05-APR-1992 (first entry)  
 DE Sequence encoded by the L1 region of the human papillomavirus (hpv)  
 DE type Ia genome.  
 KW Diagnostic reagent; vaccine; medicine; wart; tumour.  
 OS Human papillomavirus.  
 PN EP-92456-A.  
 PD 26-OCT-1983.  
 PR 01-APR-1983; 901081.  
 PR 05-APR-1982; FR-005687.  
 PA (INSP) INST PASTEUR.  
 PA (DANO) DANOS O.  
 PI Danos O, Katinka M, Yaniv M;  
 DR WPI; 83-802979/44.  
 DR N-PSDB; N30174.  
 PT DNA fragment coding for Papillomavirus antigenic proteins - and  
 PT derived immunogen, vaccine and antibody  
 PS Disclosure; Fig 1; 25pp; French.  
 CC The inventors claim DNA fragments capable of expressing, in a host,  
 CC a prod. contg. at least one antigenic determinant of papillomavirus  
 CC (PV), (see N30170-N30173). Also claimed are immunogens consisting  
 CC of at least one peptide sequence coded for by the DNA fragments (see  
 CC P30310-P30313), vaccines contg. the immunogens and antibodies raised  
 CC from them. The vaccines are useful in human and veterinary medicine  
 CC and the antibodies are useful as diagnostic reagents. The DNA  
 CC fragments are most esp. derived from the L1 region of human PV type  
 CC Ia.  
 SQ Sequence 508 AA;

Query Match 53.5%; Score 1992; DB 4; Length 508;  
 Best Local Similarity 49.4%; Pred. No. 2,946-188;  
 Matches 250; Conservative 138; Mismatches 107; Indels 11; Gaps 8;  
 Db 7 navlpaqnfkylppqptrilstdteytrtnlfyhatserlllvghplfaissn--qtv 64  
 Qy 1 MALWRPSDNTYLVPPPSVARVNTDDYVTRTSIFVHAGSSRLLVGNFYFVPVAGGNNKQ 60  
 Db 65 tipkvspnafvfrfapdnrfafgkaifnpeterlvwlgieigrqplgigtgh 124  
 Qy 61 DIPKVSAYQYRVFVQLPDPNKGFLPDNSIYNPETQRLVWACAGVEIGRGQPLGVGLSGH 120  
 Db 125 pllnklldaenptnyinhangdrqrntafakqgtqmfvgctpasgheht-sarcpcgeq 183  
 Qy 121 PFYNKLDDES-SHAATSNVSEDVNDVYKQTLQILCAPAIGEHAKGACKSRP 179  
 Db 194 vklgdcprvqmlesviedgmdndfgandfaalqgkdsdpdvqatckypdyimnh 243  
 Qy 180 LSQGDCCPPLKNTVLEGGDMVDVGYGAMDFSTLQDTKEVPLDTCQICKYPPDYQLQMSA 239

Db 244 eaygnsmffarreqmytrhfftrggsvgkavpqslyltadaeprttlattnvgtpps 303  
 Qy 240 DPGDSMFFCLRRQLFARHFNWNRAGTWGDIVPQSLYIKG-TGMRASPGSCVYSPSPSGI 296  
 Db 304 gsmvssdqlfnrswlqroqggnngicwnqlfitygdntgrtslsi--smknaasty 361  
 Qy 297 GSIVTSQSLFNKPYLWHLKAQHNGNHCWNLQFLVTVVDTTSTNLITICASTQSPVPGQY 356  
 Db 362 snaafndflrhteefdlflvqlckvkltpenlaylhtmdpnilledwqlsvsqptuple 421  
 Qy 357 DATFKQYSRHEVEYDQFIFQLCTITLTADVMSYIHSNMSSILEDWNFVPPPTTSLV 416  
 Db 422 dqyiflgsslaakcpeappetpdyqskfwevdltermseqldgplgrkfkllqsgm 481  
 Qy 417 DTYRFV-QSVAITCQDAAPENKDPYDKLKFVNDLKEKFSLDLDQYPLGRKFLVQAGL 475  
 Db 482 tqrtatsttkrtkvrvstsaakrrk 507  
 Qy 476 -RRKPTIGPKRSAPSATTSSKPAKRVVRARK 500

## RESULT 5

ID P51099 standard; Protein; 495 AA.  
 AC P51099;  
 DT 21-NOV-1991 (first entry)  
 DE Sequence of bovine papillomavirus capsid protein L1.  
 DE vaccine; papillomavirus; viral capsid protein.  
 KW Bovine papillomavirus.  
 OS Papillomavirus.  
 PN EP-133123-A.  
 PD 13-FEB-1985.  
 PR 23-JUL-1984; 401548.  
 PR 25-JUL-1983; US-517063.  
 PR 11-JUL-1984; US-629852.  
 PA (MOLE-) MOLECULAR GENETICS.  
 PI Pilacinski WP, Glassman DL, Krzyzek RA, Sadowski PL;  
 DR WPI; 85-039686/07.  
 DR N-PSDB; N50262.  
 PT prodn. of protein(s) related to papilloma virus capsid protein(s)  
 PT - by recombinant DNA techniques for vaccine prodn.  
 PS Disclosure; Fig 3; 100pp; English.  
 CC The inventors claim a recombinant DNA vector, comprising: a DNA  
 CC sequence coding for an immunoreactive and antigenic determinant of a  
 CC papillomavirus protein wherein the DNA SQ codes for L1 or L2 as  
 CC depicted in Fig 3. Using the vector, proteins related to viral  
 CC capsid proteins of papillomavirus can be obtd. by recombinant DNA  
 CC techniques and used as vaccines.  
 SQ Sequence 495 AA;

Query Match 50.0%; Score 1862; DB 3; Length 495;  
 Best Local Similarity 48.8%; Pred. No. 7,266-175;  
 Matches 245; Conservative 121; Mismatches 124; Indels 12; Gaps 10;  
 Db 1 malwqggqk-lylppptvskvscetvyrksifvhaeterlltighpyy--plsigdk- 56  
 Qy 1 MALWRPSDNTYLVPPPSVARVNTDDYVTRTSIFVHAGSSRLLVGNFYFVPVAGGNNKQ 60  
 Db 57 tvpkvsanqyrvfkiqldpnpqfalpdrthmpskervlwaivgvsvrgqlggtvtgh 116  
 Qy 61 DIPKVSAYQYRVFVQLPDPNKGFLPDNSIYNPETQRLVWACAGVEIGRGQPLGVGLSGH 120  
 Db 117 ptfalldaenlnrkvttqtdtdd-rkqtdgldakqgqlllqctpaegewtarpvcvtdr 175  
 Qy 121 PFYNKLDDES-SHAATSNVSEDVNDVYKQTLQILCAPAIGEHAKGACKSRP 179  
 Db 176 lengacppllelknkhiedgdmmeigfgaanfkeinaasksdpldignelcylpdyklmae 235  
 Qy 180 LSQGDCCPPLKNTVLEGGDMVDVGYGAMDFSTLQDTKEVPLDTCQICKYPPDYQLQMSA 239  
 Db 236 daagnsmffarqevyrvhlwtrggsekeapttdfylvknkkgdatlkipsvfhgspgs 295  
 Qy 240 DPGDSMFFCLRRQLFARHFNWNRAGTWGDIVPQSLYIKG-TGMRASPGSCVYSPSPSGI 298



```

Db 296 lvtadnqifnryelraqmngiaonnlflvtgdntrgtntltisvpsdtplteyds 355
Qy 299 IIVTSQSLFNKFPYWLHKAQHNGICWHNQLFVWVDTTRSTNLITICASTQSPVPGQYDA 358
Db 356 skfnvyrhmeeyklafilelceveltaetvshlqgimpavlenweigvqptasiledt 415
Qy 359 TKFKQISRVIEYDQFQLQCITITADVMYSIHSMNSILEDNMFVPPPTISLVDIT 418
Db 416 yryiespatkcasnvipa-kqdpayagfkwnldkqklsldlqfgrfrflagagagc- 473
Qy 419 YRFVQSVATCOKDAAPKNDPDKLKFVNDLKEKFLSDLDQYPLGRKFLVQAGLRK 478
Db 474 stv--fkrri-sqktskspak 492
Qy 479 PTIGPKRSAPSANTSSPKAR 500

RESULT
ID R88267 standard; Peptide: 215 AA.
AC R88267;
DE 12-JUN-1996 (first entry)
DE Papilloma virus major capsid protein.
KW HP-Virus 29; papilloma virus; major;
KW capsid protein; plasmid VS93-1; DSM 9133; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PR (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PR de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI: 95-383680/50.
DR N-PSDB: T03499.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PT useful for detecting papilloma virus in skin carcinoma
PS Claim 7; Fig 7; 15pp; German.
CC 20350 is 73.1% homologous to HP-Virus 29, encodes R88267 a
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS93-1 (DSM 9133). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 215 AA;

Query Match 31.7%; Score 1178; DB 16; Length 215;
Best Local Similarity 70.2%; Pred. No. 9,866-105;
Matches 151; Conservative 37; Mismatches 27; Indels 0; Gaps 0;

Db 1 grgqplvgisghpnylnkindensiahadspdrdnisvdkqtkqlcylgctppmge 60
Qy 108 GRGQPLVGLSGHPFYFNKLDDESSHAATSNVSEYDRDNVSDYKQTLCLGCAPAIGE 167
Db 61 ywkgtpcaanttpgdcpplelmtsyiqgdmvdtgygandftalqfksdvpldicqs 120
Qy 168 HWAKGTACKSRPLSQDCQCPPELKNVTLEDGDMVDVTGYGAMDFSTLQDTKCEVPLDICQS 227
Db 121 lckypdylnmaadpygdmfflrreglfarhfnragdgdkipeslylkgssgretpg 180
Qy 228 ICKYPDYLQMSADPYGDSMFFCLRRQLFARHFNWNRAGTGMGTVPQSLYIKGTGMRASPG 287
Db 181 saiyptpsgsmvteaqifnksylwagagqng 215
Qy 288 SCVSPSPSGSIVTSDSQLENKPYWLHKAQHNG 322

RESULT
ID R88273 standard; Peptide: 221 AA.
AC R88273;
DE 12-JUN-1996 (first entry)
DE Papilloma virus major capsid protein.
KW HP-Virus 15; papilloma virus; major;
KW capsid protein; plasmid VS42-1; DSM 9139; diagnosis;


```

```

KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PR (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PR de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI: 95-383680/50.
DR N-PSDB: T03505.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PT useful for detecting papilloma virus in skin carcinoma
PS Claim 7; Fig 7; 15pp; German.
CC 20350 is 73.1% homologous to HP-Virus 15, encodes R88273 a
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS42-1 (DSM 9139). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 221 AA;

Query Match 20.5%; Score 764; DB 16; Length 221;
Best Local Similarity 51.6%; Pred. No. 8,546-63;
Matches 114; Conservative 40; Mismatches 61; Indels 6; Gaps 2;

Db 1 grgqpgvgttgthpfnklrdaensserqgdttaadrmnlsfdpdkvqgmfigtptclge 60
Qy 108 GRGQPLVGLSGHPFYFNKLDDESSHAATSNVSEYDRDNVSDYKQTLCLGCAPAIGE 167
Db 61 ywkgapvcckdagmqmglcpllelksnviedgdmfdigfnlnkltfsnrsdvsldvne 120
Qy 168 HWAKGTACKSRPLSQDCQCPPELKNVTLEDGDMVDVTGYGAMDFSTLQDTKCEVPLDICQS 227
Db 121 lckypdylnmsndvysdcffcarqcyarhyfvgvvgvgsipdgavgsnkyylasa 180
Qy 228 ICKYPDYLQMSADPYGDSMFFCLRRQLFARHFNWNRAGTGMGTVPQSLYIKGTG 281
Db 181 qnslentstfpcvsqslvtadqglfnrpfwlkragghng 221
Qy 282 NRASPGSCVSPSPSGSIVTSDSQLENKPYWLHKAQHNG 322

RESULT
ID R88268 standard; Peptide: 222 AA.
AC R88268;
DE 12-JUN-1996 (first entry)
DE Papilloma virus major capsid protein.
KW HP-Virus 49; papilloma virus; major;
KW capsid protein; plasmid CR148-59; DSM 9134; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PR (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PR de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI: 95-383680/50.
DR N-PSDB: T03500.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PT useful for detecting papilloma virus in skin carcinoma
PS Claim 7; Fig 2; 15pp; German.
CC 203500 is 75% homologous to HP-Virus 49, encodes R88268 a
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid CR148-59 (DSM 9134). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 222 AA;

Query Match 20.3%; Score 754; DB 16; Length 222;
Best Local Similarity 49.6%; Pred. No. 8,626-62;
Matches 111; Conservative 52; Mismatches 50; Indels 11; Gaps 7;


```

Db	1	srghplvgstgthplfnkxkvdtenannyivtshdd-rqdtshfopkvqvmfllgcpcigse	59
QY	108	GRGQPLGVLSGHPFYKNLDDTESSHAATSNVEDYRDNVSDYKQTQLCILGCAPAIGE	167
Db	60	hwaadapccdad-rvgvkcpplvltvntedgdmvdifgfninnkttlsanksdvsddivnn	118
QY	168	HWAKGTACKSRPLSQGDCCPPLKNTVLTLEDGMVDYGMDFSTLQDTKCEVPLDICQS	227
Db	119	ickypdfkmandlygdsccffvrraqcyarhffvrggnvgdaipdaavqgdnfnlpaa	178
QY	228	ICKYPDYQMSADPYGDSMEFFCLLRQQLFARHFWRNAGTMGDTPQSL-----YIKGTG	281
Db	179	vgaqantlgssivptvsgslvstdaqflnrfpwlraqghnng	222
QY	282	M-RA-SPGCVTSPSPSGSIVTSDQLFNKPYLWLHKAQHNNG	322

RESULT	9
ID	R88269 standard; Peptide; 220 AA.
AC	R88269;
DE	12-JUN-1996 (first entry)
DE	Papilloma virus major capsid protein.
DE	HP-Virus 49; papilloma virus; major;
KW	capsid protein; plasmid VS40-7; DSM 9135; diagnosis;
KW	skin carcinomas; therapy; vaccination.
OS	Papilloma virus.
PN	DE415743-AI.
PD	09-NOV-1995.
PF	04-MAY-1994; 415743.
PR	04-MAY-1994; DE-415743.
PI	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI	de Villiers zur Hausen E, Leigh I, Shananin V, zur Hausen H;
DR	WPI; 95-383680/50.
DR	N-PSDB; T03501.
PT	DNA encoding peptide(s) of papilloma virus major capsid protein
PT	useful for detecting papilloma virus in skin carcinoma
PS	Claim 7; Fig 3; 15pp; German.
CC	T03501 is 78.5% homologous to HP-Virus 49, encodes R88269 a
CC	peptide of papilloma virus (PV) major capsid protein and is
CC	expressed by the plasmid VS40-7 (DSM 9135). The DNA is useful
CC	in diagnosis, esp. to detect pv in skin carcinomas. Major capsid
CC	proteins encoded by a PV genome contg. the DNA are useful for
CC	therapy and/or vaccination.
SQ	Sequence 220 AA;
Query Match	19.9%; Score 742; DB 16; Length 220;
Best Local Similarity	50.0%; Pred. No. 1.38e-60;
Matches 111:	Conservative 50; Mismatches 50; Indels 11; Gaps 11;

Db	1	srqgplvgstgthpflknvkxkdtensnyit-mskddrdqtsfdpkaqvqmflilgcapcige	59
QY	108	GRGQPLGVGLSGHPFXNKLDDTESSHAATSNVSEDYRDVYKQTQLCILGCAPAICE	167
Db	60	hwdaaapcdadk-gdgkcpplvlnviedgdmvdfginnkttlsanksdvsldivnn	118
QY	168	HWAKGTACKSRPLSQGDCPPELEKNTVLEDDGMVDYTGAMDFSTLQDTKCEVPLDICO	227
Db	119	ickypdflmandlvgdscffyarrecqvarhffvvgngvdrpnaaavgvgnnfmfpaa	178
QY	228	ICKYPDYQMSADPYGDSMEFFCLRRQLFARHPFWNRAGTNGDTPVQS-L-----YI--KG	279
Db	179	agaqntlgnslyvptvsgslvstdaqlnrpfwlrqaghn	220
QY	280	TGM-RASPGSCVTSPSPGSGIVTSDSGLFNKPYLWHLKHAQGN	320

RESULT	10
ID	R88271 standard; Peptide; 224 AA.
AC	R88271;
DT	12-JUN-1996 (first entry)
DE	Papilloma virus major capsid protein.
KW	HP-Virus 17; papilloma virus; major;

KW	capsid protein; plasmid VS102-4; DSM 9137; diagnosis;
KW	skin carcinomas; therapy; vaccination.
OS	Papilloma virus.
PN	DE4415743-A1.
PD	09-NOV-1995.
PF	04-MAY-1994; 415743.
PR	04-MAY-1994; DE-415743.
PI	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA	de Villiers zur Hausen E, Leigh I, Shumanin V, zur Hausen H;
DR	WFI; 95-383680/50.
DR	N-PSDB; T03503.
FT	DNA encoding peptide(s) of papilloma virus major capsid protein
PT	claim 7; Fig 5; 15pp; German.
PS	Useful for detecting papilloma virus in skin carcinoma
CC	T03503 is 79% homologous to HP-virus 17, encodes R88271 a
CC	peptide of papilloma virus (PV) major capsid protein and is
CC	expressed by the plasmid VS102-4 (DSM 9137). The DNA is useful
CC	in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC	proteins encoded by a PV genome contg. the DNA are useful for
CC	therapy and/or vaccination.
CC	Sequence 224 AA;
SQ	

```

Query Match      19.6%  Score 728;  DB 16;  Length 224;
Best Local Similarity 49.8%  Pred. No. 3.50e-59;
Matches 112;  Conservative 53;  Indels 11;  Gaps

Db      1  srgqplvgtsghplnkvrtdtens-gnyqavsgddarqntsfdpkqgmfvlgcvcpmgrr 59
      :|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY      108  GRQPLGVGLSGHPFNKLLDDESSHAANSWSEYRDVNVVDYKQTCILCCAPAIQE 167
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db      60  hwdkavceseanqqglcoppelknsuviqdgmfdigfnlnnkalsynksadvslidvyn 119
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      168  HWAKGTACKSRPLSQ-GDCPPLKTKLVLEGDGDMVDTGYCAMDFSLQDTKCEVPLDICO 226
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db      120  evckyodflmandvvgdcaiffarreqcyvarhyfrgnvgdaipdgavqgdhnylpa 179
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      227  SICKYPDYLOMSADPFGDSMFFCLREQLFARHFWRACTMGDTVP-----Q---SLYIKG 279

Db      180  qnaqqqhtlgnslpyptvsgslvtsdaqlnfrpfvlqraqgqnnq 224
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      280  -TGNRASP-GSCVYSPSPSGIVTSSQLFNKPFTWLKHAQGHNG 322
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT	11
ID	R8274 standard; Peptide: 224 AA.
AC	R8274;
DT	12-JUN-1996 (first entry)
DE	Papilloma virus major capsid protein.
KE	HP-Virus 15; papilloma virus; major;
KW	capsid protein; plasmid VS92.1; DSM 9140; diagnosis;
KW	skin carcinomas; therapy; vaccination.
OS	Papilloma virus.
PN	DE415743-A1.
PD	09-NOV-1995.
PF	04-MAY-1994; 415743.
PR	04-MAY-1994; DE-415743.
PA	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI	de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR	WFI: 95-383680/50.
DR	N-FSDB; T03506.
PT	DNA encoding peptide(s) of papilloma virus major capsid protein -
PS	useful for detecting papilloma virus in skin carcinoma
PS	Claim 7; Fig 8; 15pp; German.
CC	T03506 is 82.8% homologous to HP-Virus 15, encodes R8274 a
CC	peptide of papilloma virus (PV) major capsid protein and is
CC	expressed by the plasmid VS92-1 (DSM 9140). The DNA is useful
CC	in diagnostics, esp. to detect PV in skin carcinomas. Major capsid
CC	proteins encoded by a PV genome contg. the DNA are useful for
CC	therapy and/or vaccination.
SQ	Sequence 224 AA;

Query Match	19.5%;	Score 724;	DB 16;	Length 224;
Best Local Similarity	49.8%;	Pred. No. 8.80e-59;		

```
Matches 112; Conservative 48; Mismatches 54; Indels 11; Gaps 5;
Db 1 gqllhvgtaghlfnkvdnten-nsyqgdtstddqntsfqpkvqgmfvuqcapolce 59
Qy 108 GRGQPLGVGLSGHPFYKFKLDDTSSHAATSNVSEVDVNDVYKQTLCLGCPAIGE 167
Db 60 hwdkapvcsdknagckpplentviedgmdidfgninnkvltksedvsldivn 119
Qy 168 HWAAGTACKSRPLSQ-GDCPPLKNTVLEDGMDVDTGYGAMDFSTLQDTKCEVPLDICO 226
Db 120 etckypdfltmndvvgdcfffarreccyaryhyvrgvgvdalpeavqdknfwlpa 179
Qy 227 SICKPYDYLQMSADPYGDSMFFCLRRQLFARHFWNRAGTMDTVP-QSL-----YI--- 277
Db 180 qbtqgkdiaslyftvsgslvtsdaqlfnrpfwlraggqnnq 224
Qy 278 KGTGMRA-SPGSCVSPSPSGSIVTSDSOLFKNPKYWLHKAQGHNNG 322
RESULT 12
ID R88270 standard; Peptide; 225 AA.
AC R88270;
DE 12-JUN-1996 (first entry)
KW Papilloma virus major capsid protein.
KW HP-Virus 25; papilloma virus; major;
KW capsid protein; plasmid VS20-4; DSM 9136; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-Al.
PF 09-NOV-1995.
PR 04-MAY-1994; 415743.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI; 95-383680/50.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PS Claim 7; Fig 4; 15pp; German.
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS20-4 (DSM 9136). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 225 AA;
Query Match 19.5%; Score 724; DB 16; Length 225;
Best Local Similarity 48.9%; Pred. No. 8, 80e-59;
Matches 111; Conservative 54; Mismatches 48; Indels 14; Gaps 7;
Db 1 gqllhvgtaghlfnkvdntengtygttkdd-rqnsfapklqmfiigtcpelge 59
Qy 108 GRGQPLGVGLSGHPFYKFKLDDTSSHAATSNVSEVDVNDVYKQTLCLGCPAIGE 167
Db 60 hwdkapavcn-diqgscpplentvtyggdmadigynlnfkalgqrsvslidvde 118
Qy 168 HWAAGTACKSRPLSQGDCPPLKNTVLEDGMDVDTGYGAMDFSTLQDTKCEVPLDICO 227
Db 119 ickypdflrmqndvvgdcfffarreccyaryhfvrggkpgddipdaqidagshkneyyl 178
Qy 228 ICKYPDYLOMSADPYGDSMFFCLRRQLFARHFWNRAGTMDTVP--Q-----LYI 277
Db 179 qasdgqslnsmlyftsgslvtsdaqlfnrpfwlraggqnnq 225
Qy 278 KGTGMRA-SP-GSCVSPSPSGSIVTSDSOLFKNPKYWLHKAQGHNNG 322
RESULT 13
ID R88272 standard; Peptide; 220 AA.
AC R88272;
DE 12-JUN-1996 (first entry)
KW Papilloma virus major capsid protein.
Query Match 19.3%; Score 717; DB 16; Length 220;
Best Local Similarity 49.8%; Pred. No. 4, 42e-58;
Matches 110; Conservative 62; Mismatches 52; Indels 7; Gaps 4;
Db 1 gsgqplvgvgtgthpfnkldrsensaeerlegtsddrr-nisfdpklqgmfvigtcpelge 59
Qy 108 GRGQPLGVGLSGHPFYKFKLDDTSSHAATSNVSEVDVNDVYKQTLCLGCPAIGE 167
Db 60 ywdtapyckdagsqlgcpplleknsviedgmdidfgninnkvltksedvsldivne 119
Qy 168 HWAAGTACKSRPLSQGDCPPLKNTVLEDGMDVDTGYGAMDFSTLQDTKCEVPLDICO 227
Db 120 ickypdfltmndvvgdcfffarreccyaryhfvrggavgdllipdvqdhkyyilpan 179
Qy 228 ICKYPDYLOMSADPYGDSMFFCLRRQLFARHFWNRAGTMDTVPQSLYIKGTG--MRAS 285
RESULT 14
ID R88275 standard; Peptide; 228 AA.
AC R88275;
DE 12-JUN-1996 (first entry)
KW Papilloma virus major capsid protein.
KW HP-Virus 12; papilloma virus; major;
KW capsid protein; plasmid VS75-3; DSM 9141; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-Al.
PF 09-NOV-1995.
PR 04-MAY-1994; 415743.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI; 95-383680/50.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PS Claim 7; Fig 9; 15pp; German.
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS75-3 (DSM 9141). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 228 AA;
Query Match 19.3%; Score 720; DB 16; Length 228;
```

Search completed: Wed Mar 26 08:31:06 1997  
Job time : 108 secs.

Best Local Similarity 50.9%; Pred. No. 2.21e-58;  
Matches 116; Conservative 42; Mismatches 57; Indels 13; Gaps 6;

Db 1 grqqplvgvtsghplfnkvndaenplayraqafstddrqntsfdpkqgmfiigcapcig 60  
QY 108 GRGQPLGVGLSGHPFYNKLDDESSHAATSNV-SEQVDRDNVSDYKQQLCILGCPAIG 166  
Db 61 ehvdtgercagannengtcppiklvnsiqdgmadiygnlnfrtlgensdsvaldivn 120  
QY 167 EHWAGTACKSRPLSGDCPPLELANTVLEDGDMVDTCYGAMDFSTLQDTKCEVPLDICQ 226  
Db 121 etckypdfikmndlygdsccffarraqcyarhffvrggkagddlpagaqdagtykndfy 180  
QY 227 STCKYPDYQLQMSADPYGDSMFFCLRREQLFARHFNWRACTMGDTVP--Q-----LY 276  
Db 181 ipgasgqkqknigmsyftvsgslvssdaqlnrfpwlqraqgnng 228  
QY 277 IKGT-G-MRASPSCVYSPSPSGSIVTSDSOLFNRKPYWLHKAQGHNG 322

## RESULT 15

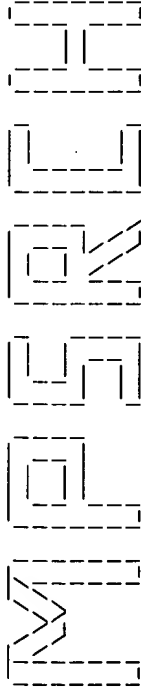
ID R32057 standard; Protein; 88 AA.  
AC R32057;  
DE 07-JUN-1993 (first entry)  
KW Rabbit serum reactive HPV16 L1 epitope 250-337.  
KW B; epitope; human; papilloma; virus; HPV-16; L1; antisera; ELISA; A1;  
KW antigenic index; flexibility; accessibility; hydrophilicity; mice;  
KW immunise; HPV16 capsid; papilloma virus like particle; VLP; diagnosis;  
KW vaccine.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Peptide 18..20  
FT /note= "Epitope 267-269"  
FT PN W09302184-A.  
PD 04-FEB-1993.  
PF 20-JUL-1992; AU0364.  
PR 19-JUL-1991; AU-007322.  
PA (CLSC-) CLS LTD.  
PA (UYQU) UNIV QUEENSLAND.  
PI Frazer I, Zhou J;  
PI WPI; 93-058783/07.  
PT Prodn. of papilloma virus-like particles - which contain L1 and  
PT L2 proteins, useful as vaccine against papilloma virus infections  
PT and in diagnosis  
PS Claim 39; Table 1; 63pp; English.  
CC The sequences given in R32054-73 are B epitopes which are derived  
CC from the human papilloma virus-16 (HPV-16) L1 protein. These  
CC epitopes were isolated by screening antisera from immunised animals  
CC against a set of overlapping HPV-16 L1 peptides by ELISA. Possible  
CC antigenic regions were identified using an antigenic index (AI) on the  
CC basis of chain flexibility, high accessibility and a high degree of  
CC hydrophilicity. A region with an AI value of over 1.5 was regarded as  
CC a predicted B epitope. Five regions identified in this manner were  
CC found to be within the 22 peptides to which major reactivity was  
CC seen with antisera from mice immunised with synthetic HPV16 capsids.  
CC These epitopes can be used to in the production of papilloma virus  
CC like particles (VLPs) for use as diagnostic agents and as components  
CC of vaccines for use with papilloma virus infections.  
SQ Sequence 88 AA;

Query Match 12.5%; Score 466; DB 6; Length 88;  
Best Local Similarity 65.9%; Pred. No. 3.33e-33;  
Matches 58; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

Db 1 ictsickypdyikmvsepygdsldffylreqmfvrhlfragatgenvpddlylkgsgst 60  
QY 224 ICQSIKCKYDYQLQMSADPYGDSMFFCLRREQLFARHFNWRACTMGDTVPQSLYIKGTGM 283  
Db 61 anlansnyftpsgsmvtsdaqifnkpy 88  
QY 284 ASPGSCVYSPSPSGSIVTSDSOLFNRKPY 311



\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Mar 27 14:05:45 1997; MasPar time 751.98 Seconds

Tabular output not generated. 893.808 Million cell updates/sec

Title: >US-08-409-122-1  
Description: (1-1524) from US08409122.seq

Perfect Score: 1524

N.A. Sequence: 1 ATGGCTTTGGGGCCCTAG.....GTGACGTGCGGAGTAA 1524  
Comp: TACCGAACACCCCGGATC.....CACATGACGCTCTTCATT

Scoring table: TABLE default

Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 630489 seqs, 220513910 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: EST-STS

1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8  
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14  
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20  
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26  
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32  
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38  
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44  
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50  
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56  
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62  
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68  
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74  
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80  
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86  
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92  
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98  
99:EST99  
EST-STS-TWO  
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104  
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109  
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114  
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119  
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124  
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129  
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134 135:EST135  
136:EST136 137:EST137 138:EST138 139:EST139 140:EST140  
141:EST141 142:EST142 143:EST143 144:EST144 145:EST145  
146:EST146 147:EST147 148:EST148 149:EST149 150:EST150  
151:EST151 152:EST152 153:EST153 154:EST154 155:EST155  
156:EST156 157:EST157 158:EST158 159:EST159 160:EST160  
161:EST161 162:EST162 163:EST163 164:EST164 165:EST165  
166:EST166 167:EST167 168:EST168 169:EST169 170:EST170 171:EST171 172:EST172 173:EST173 174:EST174 175:EST175 176:EST176 177:EST177 178:EST178 179:EST179 180:EST180 181:EST181 182:EST182 183:EST183 184:EST184 185:EST185 186:EST186 187:EST187 188:EST188 189:EST189 190:EST190 191:EST191 192:EST192 193:EST193 194:EST194 195:EST195 196:EST196 197:EST197 198:EST198 199:EST199

Database:

173:enSTS1 174:enSTS2

Statistics: Mean 11.323; Variance 1.974; scale 5.736

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
C 1	25	1.6	391 74	R10310	Yf36h09.r1 Homo sapie	9.38e-08
C 2	24	1.6	418 160	HS288304	Yf76e04.r1 Homo sapie	2.21e-06
C 3	24	1.6	418 141	N78288	Yf76e04.r1 Homo sapie	2.21e-06
C 4	24	1.6	418 70	N78288	Yf76e04.r1 Homo sapie	2.21e-06
C 5	25	1.6	442 28	H71681	Ys01c07.r1 Homo sapie	9.38e-08
C 6	23	1.5	477 9	H10719	Ym07d11.s1 Homo sapie	4.66e-05
C 7	22	1.4	162 107	T19069	f09005s Homo sapiens	8.78e-04
C 8	21	1.4	205 2	BNAF1787	Brassica campestris (	1.46e-02
C 9	21	1.4	231 93	R76852	Yf33d10.s1 Homo sapie	1.46e-02
C 10	21	1.4	243 153	W20635	mc16f08.r1 Soares mou	1.46e-02
C 11	21	1.4	243 170	MM63512	mc16f08.r1 Soares mou	1.46e-02
C 12	21	1.4	248 128	T99295	Ys63q08.r1 Homo sapie	1.46e-02
C 13	22	1.4	250 79	R26164	Yh39e12.s1 Homo sapie	8.78e-04
C 14	22	1.4	257 151	W17718	mb77f05.r1 Soares mou	8.78e-04
C 15	22	1.4	257 171	MM7189	mb77f05.r1 Soares mou	8.78e-04
C 16	21	1.4	260 30	H80328	Ys96e09.r1 Homo sapie	1.46e-02
C 17	21	1.4	263 44	HSCZOD022	H. sapiens partial CD	1.46e-02
C 18	21	1.4	267 79	R25478	Yh41h06.r1 Homo sapie	1.46e-02
C 19	22	1.4	271 48	HDM292B04B	Human aorta cDNA 5'-e	8.78e-04
C 20	21	1.4	274 44	HSCZOD012	H. sapiens partial CD	1.46e-02
C 21	21	1.4	301 95	R81312	Yj01g05.r1 Homo sapie	1.46e-02
C 22	22	1.4	332 47	HUM112B01A	Human fetal brain cDN	8.78e-04
C 23	21	1.4	356 22	H53416	Yt86h01.r1 Homo sapie	1.46e-02
C 24	21	1.4	357 46	HUM014B03A	Human fetal brain cDN	8.78e-04
C 25	22	1.4	360 94	R78356	Yf78a09.s1 Homo sapie	8.78e-04
C 26	21	1.4	363 136	HS4053YH9	H. sapiens (D12S1707)	1.46e-02
C 27	21	1.4	382 111	T37006	EST102051 Saccharomyc	1.46e-02
C 28	21	1.4	390 148	W07840	Z804a08.r1 Soares fet	1.46e-02
C 29	21	1.4	400 104	T04435	482 Arabidopsis thali	1.46e-02
C 30	22	1.4	400 133	G11249	human STS SHGC-11195	8.78e-04
C 31	22	1.4	406 26	H64976	Ys66a09.r1 Homo sapie	8.78e-04
C 32	22	1.4	407 79	R28023	Yh58g06.s1 Homo sapie	8.78e-04
C 33	21	1.4	411 160	HS313320	Za08e01.r1 Soares mel	1.46e-02
C 34	22	1.4	415 11	H16703	Ym26f12.s1 Homo sapie	8.78e-04
C 35	21	1.4	418 82	R39694	Yc97f02.s1 Homo sapie	1.46e-02
C 36	22	1.4	422 83	R40554	Yf79d03.s1 Homo sapie	8.78e-04
C 37	22	1.4	423 146	W01843	Za80a01.r1 Soares fet	8.78e-04
C 38	22	1.4	423 166	HS843329	6393 Arabidopsis thal	1.46e-02
C 39	21	1.4	430 133	T43130	Yi45h11.r1 Homo sapie	1.46e-02
C 40	21	1.4	431 91	R69853	Yq14a11.r1 Homo sapie	1.46e-02
C 41	21	1.4	441 96	R83692	20306 Arabidopsis tha	1.46e-02
C 42	21	1.4	455 157	AT26618	Yx36h11.s1 Homo sapie	8.78e-04
C 43	22	1.4	473 52	N20104	Za05f08.r1 Soares mel	1.46e-02
C 44	21	1.4	551 146	W02574	human STS WI-90599	1.46e-02
C 45	21	1.4	579 131	G07099		

ALIGNMENTS

1 R10310 391 bp mRNA EST 06-APR-1995  
LOCUS Yf36h09.r1 Homo sapiens cDNA clone 128993 5'  
DEFINITION R10310  
ACCESSION Yf36h09.r1 Homo sapiens cDNA clone 128993 5'  
NID G762266  
KEYWORDS EST  
SOURCE human clone-128993 library-Soares fetal liver spleen INFLS  
vector-pr7T3D (Pharmacia) with a modified polylinker host-DH10B  
(ampicillin resistant) primer-M13R1 Rsite1-Pac 1 Rsite2-Eco RI  
Liver and spleen from a 20 week-post conception male fetus. 1st  
strand cDNA was primed with a Pac I - oligo(dT) primer [5']









Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.

## REFERENCE

1 (bases 1 to 231)

## AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.

## TITLE

The WashU-Merck EST Project

## JOURNAL

## COMMENT

Unpublished (1995)  
Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

High quality sequence stops: 217

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

## FEATURES

Location/Qualifiers

1..231

/organism="Homo sapiens"

/clone="143923"

/notes="human"

## BASE COUNT

92 a 35 c 17 g 75 t 12 others

## ORIGIN

Query Match 1.4%; Score 21; DB 93; Length 231;

Best Local Similarity 72.7%; Pred. No. 1.46e-02;

Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 66 ataatcttcattcaataacacatgaataattataa 109

||||| | | | | | | | | | | | | | | | | |

Cp 1166 AATACATCTGCAGTAAAGTAAGTAAGTAACTGAATAA 1123

## RESULT 10

LOCUS W20635 243 bp mRNA EST 03-MAY-1996  
DEFINITION mc16f08.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 348711

## ACCESSION

W20635

GI297517

## KEYWORDS

EST.

## SOURCE

house mouse.

## ORGANISM

Mus musculus  
Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae; Murinae; Mus.

1 (bases 1 to 243)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

## TITLE

## JOURNAL

## COMMENT

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:220511

Seq primer: mob.REGA+ET

High quality sequence stop: 186.

## FEATURES

## source

NCBI gi: 1297517  
Location/Qualifiers  
1..243  
/organism="Mus musculus"  
/note="Vector: pT7f3D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7f3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo."

/clone="348711"

/dev\_stage="19 weeks"

/lab\_host="DH10B (ampicillin resistant)"

<1..>243

## BASE COUNT

95 a 51 c 37 g 60 t

## ORIGIN

Query Match 1.4%; Score 21; DB 153; Length 243;

Best Local Similarity 88.9%; Pred. No. 1.46e-02;

Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 125 aactcaagaacaccttgctttaa 151

||||| | | | | | | | | | | | | | | | | |

Cp 1379 AAGTCCAAAGAAACTTTCTCTTAA 1353

## RESULT 11

ID MM63512 standard; RNA; EST; 243 BP.

AC W20635;

DT 08-MAY-1996 (Rel. 47, Created)

DE 08-MAY-1996 (Rel. 47, Last updated, Version 1)

DE mc16f08.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 348711

DE 5'

KW EST.

OS Mus musculus (mouse)

OC Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;

OC Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae.

RN [1]

RP 1-243

RA Marra M., Hillier L., Allen M., Bowles M., Dietrich N.,

RA Dubuque T., Geisel S., Kucaba T., Lacy M., Le M., Martin J.,

RA Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K.,

RA Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R.,

RA Waterston R.

RT "The WashU-HMI Mouse EST Project";

RL Unpublished

CC Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project

CC Washington University School of Medicine 4444 Forest Park Parkway,

CC Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810

CC Email: mouseest@watson.wustl.edu This clone is available

CC royalty-free through LLNL; contact the IMAGE Consortium

CC (info@image.llnl.gov) for further information. MGI:220511 Seq

CC primer: mob.REGA+ET High quality sequence stop: 186. NCBI gi:

1297517

CC Key

Location/Qualifiers

1..243

/organism="Mus musculus"

/note="Vector: pT7f3D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7f3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo."

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

```

FT      /clone="348711"
FT      /clone_lib="Soares mouse p3NMF19.5"
FT      /dev_stage="19 weeks"
FT      /lib_host="DH10B (ampicillin resistant)"
FT      <1..243
FT      mRNA
FT      SQ Sequence 243 BP; 95 A; 51 C; 37 G; 60 T; 0 other;

Query Match      1.49; Score 21; DB 170; Length 243;
Best Local Similarity 88.99; Pred. NO. 1.46e-02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 125 aactccaaagaaccttgctcttaa 151
    ||| ||||| ||||| ||||| |||||
Cp 1379 AAGTCCAAAGAAACTTTTCCTTAA 1353

RESULT 12
LOCUS       T99295             248 bp      mRNA           EST           31-MAR-1995
DEFINITION  ye63g08.r1 Homo sapiens cDNA clone 122462 5'.
ACCESSION   T99295
NID         9749032
KEYWORDS    EST.
SOURCE      human clone=122462 library=Soares fetal liver spleen 1NFLS
vector=pMT7D (Pharmacia) with a modified polylinker host=DH10B
(vector=ampicillin resistant) primer=WL3pL Raitel-Pac 1 Rsit22-Eco RI
liver and spleen from a 20 week-post conception male fetus. 1st
strand cDNA was primed with a Fac 1 oligo (dT) primer [5'
AACGGAGAAATATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
I and cloned into the Fac 1 and Eco RI sites of the modified pMT73
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM    Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 248)
AUTHORS    Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Rulman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williams,A., Wohldman,P. and
Wilson,R.
TITLE       The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
High quality sequence stops: 237
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Location/Qualifiers
1..248
/organism="Homo sapiens"
/clone="122462"
/note="human"

BASE COUNT    94 a 62 c 39 g 53 t
ORIGIN

Query Match      1.49; Score 21; DB 128; Length 248;
Best Local Similarity 92.09; Pred. NO. 1.46e-02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 8 aatcaactgcaaatcaattcttca 32
    ||| ||||| ||||| ||||| |||||
Cp 1128 AATAACTGCAATCATATTCTCA 1104

* RESULT * 13

```









Qy	181	SGGDCPPELEKNTVLEEDGMDVFGYGAMDFSLQDTKCEVPLDQCISICKYPDYLQMSAD	240
Db	302	pygdsmffcllrreqlfarfwnragtmgdtpqgslylkgtgmpagpgcgvspsgsiv	361
Qy	241	PYGDSMFFCLLRREQLFARHFNWRAGTMGDTVPQSLYTKGTMRASPGSCVTSPPSGSIV	300
Db	362	tsdsqlfnkpywlhkaqgmngvcwhnqlfvctvtdtstpnlticasqsgvppqgydak	421
Qy	301	TSDSQLFNKPYWLHKAQGMNGICWNLQFLFVTVVDTVTRSTNLTCICASTQSPVPGQYDAK	360
Db	422	fkyqsrhveeydlqlfqlctltltadvmisyhsmnasilledwnfgvpppbttslvdtyr	481
Qy	361	FKQYSRHEEYDLQFQLCTLTLTADVMISYIHSNNSILEDWNFGVPPPTTSLVDTYR	420
Db	482	fvqsvaitcqkdaapeenkpydklfnvndlkexfldqyplgrkflvqaqlrrkt	541
Qy	421	FVQSVAITCQKDAAPENKPYDKLKFNVNDLKEKESLDDQYPLGRKFLVQAGLRKKT	480
Db	542	lqprkraspsattskpakrvvrark	568
Qy	481	lqprkraspsattsskpkakrvvrark	507
RESULT 2			
ID	VL1	HPV45	STANDARD; PRT: 539 AA.
AC	P36741		
DT	01-JUN-1994	(REL. 29, CREATED)	
DT	01-JUN-1994	(REL. 29, LAST SEQUENCE UPDATE)	
DT	01-NOV-1995	(REL. 32, LAST ANNOTATION UPDATE)	
DE	PROBABILE L1 PROTEIN.		
GN	L1.		
OS	HUMAN PAPILLOMAVIRUS TYPE 45.		
NC	VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES		
NC	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE; 94265501.		
RL	DELIUS H., HOFMANN B.;		
RL	CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).		
DR	EMBL; X74479; G397029; -.		
DR	PIR; G36566; S36566.		
SK	LATE PROTEIN.		
Qy	SEQUENCE	539 AA; 60310 MW; 727CABD1 CRC32;	
Query Match 91.3%; Score 3398; DB 9; Length 539;			
Best Local Similarity 87.3%; Pred. No. 0.00e+00;			
Matches 448; Conservative 40; Mismatches 19; Indels 6; Gaps			
Db	27	malvrpsdstvlpppsvarvvsddvartsfybagsrlltvgvpyfrvvpnsagk	86
Qy	1	NALVRSDNTVLPSPVARVNVTDYVTRTSIFYHAGSSRLTVCNPFYRV-PAGGKNK	59
Db	87	casvkvasvgyrfvalpdontkfgldstinyetarlwacvmeigrqplqigsg	146
Qy	60	QDIPKVSAYQYRVFVQLPDPNKESLPDINSIINPTQLVWACAGVIGRQPLUGVLSG	119
Db	147	hpfynkliddtesaataavicdvrdrvsvdykqtclclgcvpaigehwakctickpaq	206
Qy	120	HPFNKLDDTDESSHAATNSVSEDVRDVSVDYKQJCLIGCAPAIGEHWAGTACKSRP	179
Db	207	lqpdcpplleknltiedgmndvtgysamdftldqtckevpldicqslckypdylqmsa	266
Qy	180	LSQGDCCPPELEKNTVLEEDGMDVFGYGAMDFSLQDTKCEVPLDQCISICKYPDYLQMSA	239
Db	267	dpygdsmffcllrreqlfarfhnwragvmgdtyptdlylkgtsanmrtpgscvyspspg	326
Qy	240	DPYGDSMFFCLLRREQLFARHFNWRAGTMGDTVPQSLYTKGTG--MRASPGSCVTSPPSG	297
Db	327	siifdsqllfnkpywlhkaqgmngicwhnqlfvctvtdtstnlticasqgnvpvpsdy	386
Qy	298	SVTSDQLFNKPYWLHKAQGMNGICWNLQFLFVTVVDTVTRSTNLTCICASTQSPVPGQY	357
Db	387	pkfkfkyrsrheeydlqlfqlctltltaevmisyhsmnasilledwnfgvpppbttslvd	446

QY	358	ATKQKQSRHVEEYDLQFIQLCTITITADVMSYIHSNNSILEDWNFGVPPPTSLVD	417
Db	447	tyrfvqvavtcqkdttppekqdydklxfwtvdkkfssdlidqyplgrklvqaglr	506
QY	418	TYRFVQSVAITCOKDAAPENKDPYDKLAFVNDLKERFSLDLDQYPLGRKELVQAGLR	477
Db	507	rptgprkrpaastastastaspakrvirskk	539
QY	478	KPTIGPKRSAPSA-T--TSSKPAKRVVRANK	507

RESULT	3		
ID	VLL_HPVME	STANDARD:	PRT; 505 AA.
AC	P27964:		
DT	01-AUG-1992 (REL. 23, CREATED)		
DT	01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)		
DT	01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)		
DN	PROBABLE L1 PROTEIN.		
GN	L1		
OS	HUMAN PAPILLOMAVIRUS TYPE ME180		
OC	VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RA	MEDLINE; 91374616.		
RL	REUTER S., DELLOS H., KAHN T., HOFMANN B., ZUR HAUSEN H., SCHWARZ E.;		
RL	J. VIROL. 65:5564-5568(1991).		
DR	ENBL; M73258; --; NOT_ANNOTATED_CDS.		
DR	PIR; B40508; PIWLPR.		
DR	LATE PROTEIN.		
SK	SEQUENCE 505 AA; 56805 MW; 32FD93FE CRC32;		

Query Match	82.0%;	Score	3052;	DB	9;	Length	505;
Best Local Similarity	76.7%;	Pred.	No. 0.00e+00;				
Matches	386;	Conservative	76;	Mismatches	41;	Indels	0;
Gaps	0;						

Db	1	malwrssdmnylpppsvavkvvntddvtrtgitlyyagtslltvgbpykvpmsgggkq	60
QY	1	MALWRPSDNTVYLPSPVARVNTDDVTRTSIFVHAGSSRLITGVNPFRVPAGGNKQ	60
Db	61	dipkvsayqyviralslpdpnkfsipestlynpdqrlvwacvqveigrgqlvgylsg	120
QY	61	DIPKVSAYQYVRFVQLPDPNKFGLPDNSYINPOTRLVWACAGVEIGRGQLVGLSGH	120
Db	121	plynrlddtenspfnsknpxdsrdnsvdykqtqlcligcvepaigehwakskcspnv	180
QY	121	PFYNRLDDTSSHAATSNVSEDVRDNDVNDYKQTQLCLIGCAPAIGEHWAKTACKSRPL	180
Db	181	qpgdcpplelvnptqgdmdtdtgamfstlqetkeevpldicsvcktydylqmead	240
QY	181	QPGDCPPLELVNPTQGDMDVDTGYGAMDFSTLQTKCEVPLDICSICKYFDYLOXMSD	240
Db	241	vygdamfcrlrqlfarhfnrsgmvgdtiptelylkgtdirdspesyyapspsgm	300
QY	241	PYGDSMFCRLRRQLFARHFNRAKMGDTIPQSYIIGTGRASPECVYSPSPGSIV	300
Db	301	sdsqglfnkpywlhkeqhgngicwhnqlfltvtdtrstnftlsttstesapnlydnk	360
QY	301	TSDSGLFNKPYWLHKAQHGNGICWHNQLFVTVVTRSTNFTLSTTCASTQSPVPGGYDATK	360
Db	361	fkeylrhveeydlqiflqcltisttdmsyhtmpaillddwnfgvapppsaslvdyr	420
QY	361	FKQYSRRHVEEYDLQFIQLCTITITADVMSYIHSNNSILEDWNFGVPPPTSLVDYR	420
Db	421	ylqsaatcckdpaptkdydylgnfnwnlkekfssellqfplgrklvqagvrrpt	480
QY	421	FVQSVAITCOKDAAPENKDPYDKLAFVNDLKERFSLDLDQYPLGRKELVQAGLRKPT	480
Db	481	igprkrpatattastskhkriv	503
QY	481	IGPRKRSAPSA-TSSKPAKRVVRANK	503

RESULT	4		
--------	---	--	--

```

ID VLI_HPV39 STANDARD; PRT; 505 AA.
AC P24838;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DE PROBABLE L1 PROTEIN.
GN LI.
OS HUMAN PAPILLOMAVIRUS TYPE 39.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91135017.
RA VOLPERS C., STREECK R.E.;
RL VIROLOGY 181:419-423(1991).
DR EMBL; M62849; G463192; -.
DR PIR; H38502; P1WL39.
KW LATE PROTEIN.
SQ SEQUENCE 505 AA; 56604 MW; F07CFCE CRC32;

Query Match 81.4%; Score 3030; DB 9; Length 505;
Best Local Similarity 76.7%; Pred. No. 0.00e+00;
Matches 389; Conservative 76; Mismatches 40; Indels 2; Gaps 2;

Db 1 namrdsmsvlylppsvakvntddytrtnlyyyagsslltvghpyfkgmnggrtk 60
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 1 MALWRPSDNTVYLPPLPPSVARVNTDDYVTRTSIFYHAGSSRLLVGNPYFRVPAGGNKQ 60
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 dipkvsaygyrvfrvrlpdpnkfslpdslynpetgrlrvwacvgevgrgplvglsgh 120
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 61 DIPKVSAYQYRVFRVQLPDPNKFGLPDSIYNPETQRLVWACAGVEIGRGQPLVGLSGH 120
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 plnvrqddtens-pfssttnkdsdnvsvdykqlcligcvpaigehgwkgackpnnv 179
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 121 PFYNKLDDESSHAATSNVSEDVRDVSVDYKQTLCLGCPALGCAIGERHAWGTACKSRPL 180
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 180 stgdcpllelvntpiedgmdltyggandfgalgetksevpdlcqsckypdyqlmsad 239
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 181 SQGDCPPELKNLTVEGDGMDVDTGCGAMDFSTLQDTKCEVPLDICQICKYDYLQMSAD 240
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 240 vvgdsmfclredlfarfhnrgmvgdipadqilkgtdiranpgsvvcpssgm 299
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 241 PYGDSMFCLAREQLFAHFHWRNAGTMDTVPOSLYIKGTGMRASPGSCVYSPSPSGSIV 300
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 300 tsdsglfnkpywlhkaqngnngicwhnglfltvvdttrstnftstlessipstypsk 359
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 301 TSDQLFNKPYWLHKAQNGNNGICWHNGLFVTVVDTTRSTNLITCASQSPVPGGYDATK 360
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 360 fkeytrhveeydlqfqlctvtltdvmsyihmssildnwfavappssaslvdyr 419
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 361 FKQYSRRHVEEYDLQFIFQLCTITITADVMSYIHSNMSSILEDNWFVGPPTTSLVDYR 420
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 420 ylgaaaitcdapapekdpdydglkfnvdrkelsldqfqlgrkflilqarvrrpt 479
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 421 FVQSVAITCQDAAPAEKNKDPYDLKFWNVDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 480 lgpkrpaastsses-akhrkrvsk 505
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 481 IGPKRSPATSATSSKPAKRVRR 507
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 5
ID VLI_HPV10 STANDARD; PRT; 531 AA.
AC P36732;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DE PROBABLE L1 PROTEIN.
GN LI.
OS HUMAN PAPILLOMAVIRUS TYPE 10.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94265501.

Query Match 73.2%; Score 2723; DB 9; Length 532;
Best Local Similarity 68.8%; Pred. No. 0.00e+00;
Matches 349; Conservative 91; Mismatches 60; Indels 7; Gaps 4;

```

```

RA DELIUS H., HOFMANN B.;
RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-29 IS THE INITIATOR.
DR EMBL; X74465; G396908; -.
DR PIR; S36537; S36537.
KW LATE PROTEIN.
SQ SEQUENCE 531 AA; 59011 MW; 5A30125F CRC32;

Query Match 73.9%; Score 2751; DB 9; Length 531;
Best Local Similarity 67.9%; Pred. No. 0.00e+00;
Matches 342; Conservative 103; Mismatches 57; Indels 2; Gaps 2;

Db 29 malwrsdnlvlylppsvkylstddvtrtnlyyyagtslltvghpyfipkssnnkv 88
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 1 MALWRPSDNTVYLPPLPPSVARVNTDDYVTRTSIFYHAGSSRLLVGNPYFRVPAGGNKQ 60
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 89 dvpkvsaygyrvfrvrlpdpnkfslpdaerlynpdaerlywactvevgrgplvglsgh 148
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 61 DIPKVSAYQYRVFRVQLPDPNKFGLPDSIYNPETQRLVWACAGVEIGRGQPLVGLSGH 120
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 149 plvynkledtenslahpiggdsrdnlsvdnkqtlcligctppmgehgwkgtpcrpp- 207
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 121 PFYNKLDDESSHAATSNVSEDVRDVSVDYKQTLCLGCPALGCAIGERHAWGTACKSRPL 180
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 208 agqdcplleltsipqgdmdvtyggamdfalqlnksdvpdlcqsckypdyqlmsae 267
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 181 SQGDCPPELKNLTVEGDGMDVDTGCGAMDFSTLQDTKCEVPLDICQICKYDYLQMSAD 240
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 268 pygdsmfylrredlfarfhnrasavgdalpdtdfllksnggrdvgsavyspssgm 327
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 241 PYGDSMFCLAREQLFAHFHWRNAGTMDTVPOSLYIKGTGMRASPGSCVYSPSPSGSIV 300
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 328 tseaglfknpywlhkaqngnngicwhnglfltvvdttrstnmlcvpseapattydatk 387
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 301 TSDQLFNKPYWLHKAQNGNNGICWHNGLFVTVVDTTRSTNLITCASQSPVPGGYDATK 360
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 388 fkeytrhveeydlqfqlctvtltdvmsyihmssilednwfngitlpstledtyr 447
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 361 FKQYSRRHVEEYDLQFIFQLCTITITADVMSYIHSNMSSILEDNWFVGPPTTSLVDYR 420
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 448 flssaatcqdtpptekqdyaklnfwdvdlkdrfldslsqfplgrkflilqgvrrsa 507
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 421 FVQSVAITCQDAAPAEKNKDPYDLKFWNVDLKEKFSLDLDQYPLGRKFLVQAGLRKPT 480
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 508 vsrkrpatsatgst-aakrtrtk 530
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 481 IGPKRSPATSATSSKPAKRVRR 504
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

RESULT 6
ID VLI_HPV03 STANDARD; PRT; 532 AA.
AC P36731;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DE PROBABLE L1 PROTEIN.
GN LI.
OS HUMAN PAPILLOMAVIRUS TYPE 3.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94265501.
RA DELIUS H., HOFMANN B.;
RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-29 IS THE INITIATOR.
DR EMBL; X74462; G397012; -.
DR PIR; S36554; S36554.
KW LATE PROTEIN.
SQ SEQUENCE 532 AA; 59194 MW; EFADFC13 CRC32;

Query Match 73.2%; Score 2723; DB 9; Length 532;
Best Local Similarity 68.8%; Pred. No. 0.00e+00;
Matches 349; Conservative 91; Mismatches 60; Indels 7; Gaps 4;

```

```
Db 29 malwrsdnlvlpptvskvslatdyvtrtniyyagssrllltvghpfaipakssnsk 88
QY 1 MALWRPNDNTVYLPVPPSVARVWNTDDYVTRTSIFVHAGSSRLLTGVPYFRVPAGGNKQ 60
Db 89 dipkvaafyrvfvrhlpdpnkfgldpazfynpdaerlvwactvgvgrgplvglgsh 148
QY 61 DIPKVASAYQYRVFRVQLPDPNKFGLPDNSIYNPETQRLVWACAGVEIGRGQPLVGLSGH 120
Db 149 plnykliddtenshatndiqkdrdnisvndkqglclvictpctpmehgwkgctckqna- 207
QY 121 PFYNKLDDESSHAATSNVSEDVRDNVSDYKQTLQILCAPAIGEHAKGTACKSRPL 180
Db 208 spdcplleltatpdaqndvctgandfnqlsnksdvpldctctkypdylgmaae 267
QY 181 SQGDCPPELKNVLEQDFARHFNWNRAGTMDTVPOSUYIKG--TGMRASPGSCVYSPSPSG 240
Db 268 pygdsmfirhgeeyldqfqlckltltdmayihlmaslledwnfgltltpstaled 327
QY 241 PYGDSMFFCLRRQELFARHFNWNRAGTMDTVPOSUYIKG--TGMRASPGSCVYSPSPSG 297
Db 328 smvtsetqlfknkpylrragghngicwgnlfvtrvtdtrstnmtlcvstetst--yd 385
QY 298 SIYTSQSOLFKNKPYLHKAQHNGICWENOLFVTVVDTTSTRNLICASTQSPVPGQYD 357
Db 386 atkfeylrrhgeeyldqfqlckltltdmayihlmaslledwnfgltltpstaled 445
QY 358 ATKFKQYSHVEYDLQFIFQLCTITLTADVMSTYHNSNILEDNWFGVPPPTTSLVD 417
Db 446 tyrltssaitcqdapptekqpyaklnfwdvdkrdfrldsqfplgrkflmglgvgt 505
QY 418 TYRFVQSVAITCQKDAAPAENKDPYDKLFKFWNVDLKEFSLDLDQYPLGRKFLVQAGLR 477
Db 506 rssiisvkrtsa--tttsrtaaaakrtk 531
QY 478 KPIGRKRSAPSATTSSKPAKRVR 504
RESULT 7
ID VLI_HPV26 STANDARD; PRT; 503 AA.
AC P36735;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DE PROBABLE L1 PROTEIN.
GN L1.
OS HUMAN PAPILLOMAVIRUS TYPE 26.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RX MEDLINE; 94265501.
RA DELIUS H., HOFMANN B.;
RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).
RN [2]
RP SEQUENCE OF 309-352 FROM N.A.
RX MEDLINE 94407963.
RA CHAN S.Y., BERNARD H.D., ONG C.K., CHAN S.P., BIRGIT H., DELIUS H.;
RL EMBL; X74474; G396960; -.
DR EMBL; M96279; G393156; -.
KW LATE PROTEIN.
SQ SEQUENCE 503 AA; 56328 MW; 5E31786A CRC32;
Query Match 72.3%; Score 2690; DB 9; Length 503;
Best Local Similarity 67.3%; Pred. No. 0.00e+00;
Matches 337; Conservative 93; Mismatches 68; Indels 3; Gaps 3;
Db 1 malwrsdnlvlpptvskvslatdyvtrtniyyagssrllltvghpfaipakssnsk 88
QY 1 MALWRPNDNTVYLPVPPSVARVWNTDDYVTRTSIFVHAGSSRLLTGVPYFRVPAGGNKQ 60
Db 60 eipkvaafyrvfvrhlpdpnkfgldpazfynpdaerlvwactvgvgrgplvglgsh 119
QY 61 DIPKVASAYQYRVFRVQLPDPNKFGLPDNSIYNPETQRLVWACAGVEIGRGQPLVGLSGH 120
Db 120 plfnkliddtenshatndiqkdrdnisvndkqglclvictpctpmehgwkgctckqna- 179
QY 121 PFYNKLDDESSHAATSNVSEDVRDNVSDYKQTLQILCAPAIGEHAKGTACKSRPL 180
```



```
RESULT 11
ID VLI_HPV1 STANDARD; PRT; 501 AA.
AC P22163;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
GN L1.
OS RHESUS PAPILLOMAVIRUS TYPE 1 (HPV 1).
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 91303675.
RX LUNGU O., CRUM C.P., SILVERSTEIN S.J.;
RL J. VIROL. 65:4216-4225(1991).
DR EMBL; M62877; -. NOT_ANNOTATED_CDS.
KW PIR; G40415; P1WL51.
KW LATE PROTEIN.
SQ SEQUENCE 504 AA; 56314 MW; E5FA804 CRC32;

Query Match 71.0%; Score 2643; DB 9; Length 504;
Best Local Similarity 64.0%; Pred. No. 0.00e+00;
Matches 323; Conservative 112; Mismatches 67; Indels 3; Gaps 3;

Db 1 malwrtndskvylppapvsvrlynteeeylirtgyllyagssrlitlghpyfpipkte-tra 59
QY 1 MALWRPSDNTVYLPSPVARVNTDDYVTRTSIFYHAGSSRLLTGVNPFYRVPAGGNKQ 60
Db 60 alpksafgyrvfrvqlpdpnkfjlpdpnlynpdtdrlvwgcgvvegrgqplvglsgh 119
QY 61 DIPKVSAYQYRVFVQLPDPNKFGLPDNSIYNPETQRLVWACAGVEIGRQPLVGLSGH 120
Db 120 plfnkydtenariangnaqqdvrdntsvdnkqtlcligcappigehwigittcntpv 179
QY 121 PFYNKLDDESSHAATSNVSEDVRDVSVDYKQTLCLIGCAPAIGERWAKGTACKSRPL 180
Db 180 ppgdcppelelvsviqdgmldtfgamdfaalqatksdvpldisqgvckypdylkmsad 239
QY 181 SQGDCPPELEKNTVLEDGDMVDYGYGAMDFSTLQDTKCEVPLDICQICKYDYLQMSAD 240
Db 240 tygnsmfhlirreqifafahyvkilvggedipndvvyikgsgngdrpdesiyvatpsgm 299
QY 241 PYGDSMFCLRRLEQFAHFHWNAGTGMGTVPQSLYIKGTGM-RASPGSCVYSPSPSGSI 299
Db 300 itdsqifnkpwlhrragbhnglcwnnqlfctcvdtrstnltstat-aavspftps 358
QY 300 VTSQCLFNKPWLHKAQHNGICWHNQLFVTVVDTTRSTNLICASTQSPVPGYDAT 359
Db 359 nfkyvrbgeveylqfclctilttemvaylhtmdotilegnvfgltlppasleday 418
QY 360 KFKYSRHVEYDLQFQLCTITLTADVMSYIHSNNSILEDWNFGVPPPTSLVDYI 419
Db 419 fvrnaatcqdtpqkpkplagkyfcdvdlkerfeldldqfalgrkflvgvqrkp 478
QY 420 RFVQSVAITCQKDAPAKENKDPYDKLAFWVDLKERFSLDLDQIFLGRKFLVQGLRRKP 479
Db 479 rpglkrpassasassssssakkrvk 503
QY 480 TIGPKRSAPSATTSKPAKRVVR 504

RESULT 13
ID VLI_HPV31 STANDARD; PRT; 504 AA.
AC P17388;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-AUG-1990 (REL. 15, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
GN L1.
OS HUMAN PAPILLOMAVIRUS TYPE 31.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 89299478.
RX GOLDSBOROUGH M.D., DISILVESTRE D., TEMPLE G.F., LORINCZ A.T.;
RL VIROLOGY 171:306-311(1989).
DR EMBL; J04353; G459922; -.
DR PIR; G32444; P1WL31.
KW LATE PROTEIN.
SQ SEQUENCE 504 AA; 56352 MW; A8A44931 CRC32;

Query Match 70.5%; Score 2624; DB 9; Length 504;
Best Local Similarity 65.1%; Pred. No. 0.00e+00;

RESULT 12
ID VLI_HPV51 STANDARD; PRT; 504 AA.
AC P26536;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
GN L1.
OS HUMAN PAPILLOMAVIRUS TYPE 51.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.

Db 1 msnwrpsdkvylppvsvkvstdevrtsiyfhagssrllavghpyyavkkn-nkv 59
QY 1 MALWRPSDNTVYLPSPVARVNTDDYVTRTSIFYHAGSSRLLTGVNPFYRVPAGGNKQ 60
Db 60 svpkvsglgyrvfrvlpdpnkfjlpdpnlynpdtdrlvwacigvvegrgqplvgtsgh 119
QY 61 DIPKVSAYQYRVFVQLPDPNKFGLPDNSIYNPETQRLVWACAGVEIGRQPLVGLSGH 120
Db 120 pllnkliddtengpkvaggqgdreecvndyktqlclmclgckpveghwknpcctt-g- 177
QY 121 PFYNKLDDESSHAATSNVSEDVRDVSVDYKQTLCLIGCAPAIGERWAKGTACKSRPL 180
Db 178 aegdcplalelvsviqdgmldtfgamdfaalqatksdvpldisqgvckypdylkmsad 237
QY 181 SQGDCPPELEKNTVLEDGDMVDYGYGAMDFSTLQDTKCEVPLDICQICKYDYLQMSAD 240
Db 238 pygdsiflylrreemfhrhfragtmcdsvddylkgsnkvlaashfyvtpstsmv 297
QY 241 PYGDSMFCLRRLEQFAHFHWNAGTGMGTVPQSLYIKGTGMASPGSCVYSPSPSGSI 300
Db 298 tsdaqifnkpwlqagbhnglcwnnqlfctcvdtrstnltstatstastvttpynnes 357
QY 301 TSDQCLFNKPWLHKAQHNGICWHNQLFVTVVDTTRSTNLICASTQSPVPGYDATK 360
Db 358 fkeylthveefdlqfclctilttemvaylhtmdotilegnvfgltlppasleday 417
QY 361 KFKYSRHVEYDLQFQLCTITLTADVMSYIHSNNSILEDWNFGVPPPTSLVDYI 420
Db 418 fvtssaatcqpappkpkedplaktyfvevdikefseadldqfplgrkflqagmrarpt 477
QY 421 FVQSVAITCQKDAPAKENKDPYDKLAFWVDLKERFSLDLDQIFLGRKFLVQGLRRKP 480
Db 478 lrapkrtt-sstssssprkrkrk 500
QY 481 IGPKRSPAPSATTSKPAKRVVR 504

RESULT 12
ID VLI_HPV51 STANDARD; PRT; 504 AA.
AC P26536;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
GN L1.
OS HUMAN PAPILLOMAVIRUS TYPE 51.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
```

[illegible]

```

RESULT 15
ID VL1_HPV2A STANDARD; PRT; 510 AA.
AC P25486;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DE 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
GN L1.
OS HUMAN PAPILLOMAVIRUS TYPE 2A.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE: 9118699.
RA HRSCB-BEHNAM A., DELIUS H., DE VILLIERS E.M.;
RL VIRUS RES. 18:81-98(1990).
DR ENBL; X55964; -; NOT_ANNOTATED_CDS.
DR PIR; S15620; S15620.
KW LATE PROTEIN.
SQ SEQUENCE 510 AA; 57193 MW; EB7FCF70 CRC32;

Query Match 69.78; Score 2592; DB 9; Length 510;
Best Local Similarity 64.18; Pred. No. 0.00e+00;
Matches 323; Conservative 111; Mismatches 59; Indels 11; Gaps 9;

Db 17 malwpnesxvllppcpvsavistdyvttrtnvyhggsrllltvcghpyysikksn-nkv 75
|||||...|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 1 MALWRPSDNVTYVLPSPSVARVNWDYVTRTSFYHAGSRLLTVGNFRVPAGGKNQ 60

Db 76 avpkvsqyrvfhnvlpdnpkfgldpadlpydtdtllwacvgvgevrgqplgvqvsg 135
|||||...|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 61 DIPKVSAYOYRVRVOLPDNPKGLPDSNTYNPTORLWVACAGVGIRGOPIGVGLSGH 120
|||||...|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

[illegible]

Db	193	aegdcplqftntiedgmvetgfgaldfatlqsnksdvpldictatckypdykmaae	252
QY	181	SQGDOPPLELKNVLEDDGMDVTGYGAMDFSTLQDTKCEVPLDQCSTCKYDPYLQMSAD	240
Db	253	pygdsmfslrrqcmftrhfqlagkmgdtipdellykatsv-otpsghvytstpsgsmv	311
QY	241	PYGDSMEFCLRRQLFARHFNWAGTWGDIVPQSLYIKGTGNRASPSGVSPSPGSIV	300
Db	312	sseqqlfnkpylrragghnngmcwgnrvfltvdttrstnvsica-teas-dtnykatn	369
QY	301	TSQSOLFNFILHKAQHNGGICWHNQLFVIVDTIRSNLTICASTQSPVPCQYDAIK	360
Db	370	fkelylrmeeydlqfqlckititlpelmayihmdpqilledwnfgvpppppsasldqdyr	429
QY	361	FQFSRVEEYDQFIFQCTITLADVMSYIHSNMSSILEDNWEGVPPPTTSLVDYTR	420
Db	430	ylsqgaitcoktpptptdpyasltfdvdlsefsmdldqfplgrkflqrga--mpt	487
QY	421	FVQSVAITCOKDAAPAKENKDPYDKFNWVNDLXKESLDDQYPLGRKFLVQAGLRKPT	480
Db	488	vs-rkraavsgttptt-skkrivr	509
QY	481	IGPKRSAPSATTSSKPAKRVR	504

Search completed: Wed Mar 26 08:27:53 1997  
Job time : 104 secs.











































```

RESULT 2
ID VLL_HP45 STANDARD; PRT; 539 AA.
AC P36741;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PROBABLE L1 PROTEIN.
LN
GS
ON HUMAN PAPILLOMAVIRUS TYPE 45.
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; PAPOVAVIRIDAE; PAPILLOMAVIRUSES.
OX
RX SEQUENCE FROM N.A.
RX MEDLINE; 94265501.
RA DELIUS H., HOFMANN B.:
RL CORR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).
DR EMBL; X7479; G397029; -.
DR PIR; S36566; S36566.
KW LATE PROTEIN.
SQ SEQUENCE 539 AA; 60310 MW; 727CA8D1 CRC32;
(1)
Query Match 64.1%; Score 4888; DB 9; Length 539;
Best Local Similarity 54.4%; Pred. No. 0.00e+00;
Matches 814; Conservative 331; Mismatches 341; Indels 9; Gaps 5;

Db Dd
559 gntgvtncngcnathgngarcaraytggnnaargnacnyntngayaaacngcncar 618
Dt
559 gntgvtncngcnathgngarcaraytggnnaargnacnyntngayaaacngcncar 618
QY
478 GGCTGTGCCCTCTATTGGGAACACATGGCTAAAGCACTGCTTGTAAATCGGTCTCT 537
Qt
478 GGCTGTGCCCTCTATTGGGAACACATGGCTAAAGCACTGCTTGTAAATCGGTCTCT 537
G C A P A I G E H W A K G T A C K S R P

Db Dd
619 lqpcgdcpcpllelknntlelqdgdl 678
Dt
619 lqpcgdcpcpllelknntlelqdgdl 678
QY
538 TTATCATCGGCGATGGCCCTTTAGAACCTTAAGAACACAGTTTGAAGATGTGAT 597
Qt
538 TTATCATCGGCGATGGCCCTTTAGAACCTTAAGAACACAGTTTGAAGATGTGAT 597
L S Q G D C P P L E L K N T V L E D G D

Db Dd
679 atggtngaycngnctayggngcnatggaytywnacnyntngayaaacngcncar 738
Dt
679 atggtngaycngnctayggngcnatggaytywnacnyntngayaaacngcncar 738
QY
598 ATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657
Qt
598 ATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657
M V D T G I G A M D F S T L Q D T K C E

Db Dd
739 gtccnctngayathgycarwsnathgycarwsnathgycarwsnathgycarwsnath 798
Dt
739 gtccnctngayathgycarwsnathgycarwsnathgycarwsnathgycarwsnath 798
QY
658 GTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717
Qt
658 GTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717
V P L D I C Q S I C K Y P D Y L Q M S A

Db Dd
799 gaycncatayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 858
Dt
799 gaycncatayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 858
QY
718 GATCCTATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777
Qt
718 GATCCTATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777
D P Y G D S M F F C L R R E Q L F A R H

Db Dd
859 fwnravgvmgdtpvptdllylkg 918
Dt
859 fwnravgvmgdtpvptdllylkg 918
QY
778 TTTTGAATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837
Qt
778 TTTTGAATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837
F W N R A G T M G D T V P Q S L Y I K G

Db Dd
919 acnwsnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 978
Dt
919 acnwsnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 978
QY
838 AC-AG-G---TATCGGCTCACCTGGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 891
Qt
838 AC-AG-G---TATCGGCTCACCTGGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 891
T X X M R A S P G S C V Y S P S S G

Db Dd
979 wsnathgcnatggngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 1038
Dt
979 wsnathgcnatggngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 1038
QY
892 TCTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 951
Qt
892 TCTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 951
S I V T S D S Q L F N K P Y W L H K A Q

Db Dd
1039 gncnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 1098
Dt
1039 gncnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 1098
QY
952 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011
Qt
952 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011
G H N N G I C W H N Q L F V T V D T

Db Dd
1099 mgnwsnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 1158
Dt
1099 mgnwsnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 1158
QY
1012 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071
Qt
1012 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071
R S T N L T I C A S T Q S P V P G Q Y D

Db Dd
1159 cncnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 1218
Dt
1159 cncnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 1218
QY
1072 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1131
Qt
1072 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1131
A T K F K Q Y S R H V E E Y D L Q F I F

Db Dd
1219 caryntngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 1278
Dt
1219 caryntngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnatggngaycngnctayggngcnat 1278
QY
1132 CAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191
Qt
1132 CAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191
Q L C T I T L T A D V M S Y I H S M N S

Db Dd
s i l e n w n f g v p p p t t s l v d

```





































181 GATATCTCAGTTCGACCAATACAGATTCGGTGCAGTTCCTGACCCA 240  
 I P K V S A Y Q Y R F V Q L P P  
 238 aayaaaygnnhtygtggaayaaayntntatyaacnngayngayntnng 297  
 238 aayaaaygnnhtygtggaayaaayntntatyaacnngayngayntnng 297  
 241 AATAAATGGTTCACCTGATATATATATATATATATATATATATAT 300  
 N K F G L P D N S I Y N P E T Q R L V W  
 298 gcnvgtng 357  
 301 GCGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360  
 A C A G V E I G R G Q P L G V G L S G H  
 358 ccnyntiaaayaaayaaayaaayaaayaaayaaayaaayaaayaaay 417  
 361 CCATTTTATAAATATAGATGACACTGAAAGTTCCTGAGGAGGAGGAG 420  
 P F Y N K L D D T E S S H A A T S N V S  
 418 garygng 477  
 421 GAGGACGTTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480  
 E D V R D N V S V D Y K Q T Q L C I L G  
 478 gng 537  
 481 TGTCGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
 C A P A I G E H W A K G T A C K S R P L  
 538 ccng 597  
 541 TCACAGGCGATGCGCCCTTTAGAACCTAAGAACACAGCTTTGGA 600  
 S Q G D C P L E L K N T V L E D G D M  
 598 athgagng 657  
 601 GTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 V D T G Y G A M D F S T L Q D T K C E V  
 658 ccnyntiaaayaaayaaayaaayaaayaaayaaayaaayaaayaaay 717  
 661 CCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720  
 P L D I C Q S I C K Y P D Y L Q M S A D  
 718 acntayng 777  
 721 CCGTATGAGGATGATGATGATGATGATGATGATGATGATGATGAT 780  
 P Y G D S M F F C L R R E Q L F A R H F  
 778 tayaayaaayaaayaaayaaayaaayaaayaaayaaayaaayaaay 837  
 781 TGGATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840  
 W N R A G G T M G D T V P Q S L Y I K G T  
 838 gng 897  
 841 GGTATG--CGTCGTCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 897  
 G X R A S P G S C V Y S P S P S G S I  
 898 i t s d s q i f n k p y w l h r a q g h  
 athcnwng

898 GTTACCTCTGACTCCACGTTGTTTAAATAAACCATATATGTTACATAAGGCACAGGTCAT 957  
 V T S D S Q L F N K P Y W L H K A Q G H  
 958 aayaaaygnnhtygtggaayaaayntntatyaacnngayngayntnng 1017  
 958 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1017  
 N N G I C W H N Q L F V T V V D T T R S  
 1018 acnaayaaayaaayaaayaaayaaayaaayaaayaaayaaayaaay 1074  
 1018 ACCAATTAACCAATATGCTTACACAGCTCTCTGCTGCTGCTGCTG 1077  
 T N L T I C A S T V P V P G Q Y D A T  
 1075 aaytyaaayaaayaaayaaayaaayaaayaaayaaayaaayaaay 1134  
 1075 K F K Q Y S R H V E E Y D L Q F I F Q L  
 1135 tgyaaayaaayaaayaaayaaayaaayaaayaaayaaayaaayaaay 1194  
 1135 TGTAT 1197  
 C T I T L T A D V M S Y I H S M N S S I  
 1195 ytnarcarcarcarcarcarcarcarcarcarcarcarcarcarcar 1254  
 1195 TTAGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 1257  
 L E D W N F G V P P P T T S L V D T Y  
 1255 ngnttyng 1314  
 1255 R F V Q S V A I T C Q K D A A P A E N K  
 1315 gaycnyng 1374  
 1315 GATCCCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1377  
 D P Y D K L K F W N V D L K E K F S L D  
 1375 ytnarcarcarcarcarcarcarcarcarcarcarcarcarcarcar 1434  
 1375 TTAGATCAATATCCCTTGGAGCTGATGATGATGATGATGATGATG 1437  
 L D Q Y P L G R K F L V Q A G L R R K P  
 1435 ngncng 1494  
 1435 ACCATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1497  
 T I G P R K R S A P S A T T S S K P A K  
 1495 ngnaarmngngt 1505  
 1498 CGTGTGCTGT 1508  
 R V R  
 RESULT 11  
 ENTRY #type complete  
 TITLE late protein - human papillomavirus type 53  
 ORGANISM formal name human papillomavirus type 53  
 DATE 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 12-Apr-1995  
 ACCESSIONS S36531  
 REFERENCE S36469  
 #authors Dellus, H.; Hofmann, B.

















\*\*\*\*\*

W A R E L D

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Mar 27 13:44:03 1997; MasPar time 1084.04 Seconds  
Tabular output not generated. 1157.881 Million cell updates/sec

Title: >US-08-409-122-1  
Description: (1-1524) from US08409122.seq  
Perfect Score: 1524  
N.A. Sequence: 1 ATGGCTTTGTGGCGGCTAG.....GTGTACGTGCAGGAAGTAA 1524  
Comp: TACCGAAACACCGCGGATC.....CACATGACGCTCTTCATT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 279077 seqs, 411808665 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: emb1-new5  
1.BCT 2.FUN 3:INV1 4:INV2 5:ORG 6:MAM 7:VRT 8:PLN 9:PRI  
10:PRO1 11:PRO2 12:ROD 13:SYN 14:UNC 15:VIR  
Database: genbank94

16:BCT1 17:BCT2 18:BCT3 19:BCT4 20:BCT5 21:BCT6 22:BCT7  
23:BCT8 24:INV1 25:INV2 26:INV3 28:INV4 29:INV5  
30:INV6 31:INV7 32:MAM1 33:MAM2 34:MAM3 35:VRT1 36:VRT2  
37:VRT3 38:PAT1 39:PAT2 40:PAT3 41:PHG 42:PLN1 43:PLN2  
44:PLN3 45:PLN4 46:PLN5 47:PLN6 48:PLN7 49:PLN8 50:PRI1  
51:PRI2 52:PRI3 53:PRI4 54:PRI5 55:PRI6 56:PRI7 57:PRI8  
58:PRI9 59:PRI10 60:PRI11 61:PRI12 62:PRI13 63:ROD1  
64:ROD2 65:ROD3 66:ROD4 67:ROD5 68:ROD6 69:ROD7 70:ROD8  
71:SYR 72:SYN 73:UNA 74:VRL1 75:VRL2 76:VRL3 77:VRL4  
78:VRL5 79:VRL6 80:VRL7 81:VRL8

Database: genbank-news

82:BCT 83:INV1 84:INV2 85:MAM 86:VRT 87:PAT 88:PHG

89:PLN 90:PRI 91:ROD 92:STR 93:SYN 94:UNA 95:VRL

Database: u-emb146\_94

96:part1

Statistics: Mean 11.592; Variance 4.289; scale 2.703

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1484	97.4	7857	80	PAPHPV18 Human papillomavirus	0.00e+00
2	922	60.5	7858	79	HPV45 Human papillomavirus	0.00e+00
3	746	49.0	7896	79	HPV59VG Human papilloma virus	0.00e+00

	4	740	48.6	1855	79	HPV11	Human Papilloma Virus	0.00e+00
5	724	47.3	7905	79	HPU21941	Human papillomavirus	0.00e+00	
6	721	47.3	7833	80	PRT39	Human papillomavirus	0.00e+00	
7	718	47.1	6042	59	HOMHPME18	Human cellular DNA/Hu	0.00e+00	
8	535	35.1	7855	79	HPV26	Human papillomavirus	0.00e+00	
9	532	34.9	7916	79	HPU31784	Human papillomavirus	0.00e+00	
10	527	34.6	2920	15	HP37217	Human papillomavirus	0.00e+00	
11	527	34.6	2920	79	HPU37217	Human papillomavirus	0.00e+00	
12	527	34.6	7852	79	HPV30	Human papillomavirus	0.00e+00	
13	524	34.4	1517	40	113522	Sequence 1 from paten	0.00e+00	
14	525	34.4	7759	79	HPU37488	Human papillomavirus	0.00e+00	
15	515	33.8	1484	15	HP34168	Human papillomavirus	0.00e+00	
16	515	33.8	1484	79	HPU34185	Human papillomavirus	0.00e+00	
17	515	33.8	1484	15	HP34171	Human papillomavirus	0.00e+00	
18	515	33.8	1484	15	HP34181	Human papillomavirus	0.00e+00	
19	515	33.8	1484	79	HPU34181	Human papillomavirus	0.00e+00	
20	515	33.8	1484	15	HP34185	Human papillomavirus	0.00e+00	
21	515	33.8	1484	79	HPU34171	Human papillomavirus	0.00e+00	
22	515	33.8	1484	79	HPU34184	Human papillomavirus	0.00e+00	
23	515	33.8	1484	79	HPU34168	Human papillomavirus	0.00e+00	
24	515	33.8	1484	15	HP34169	Human papillomavirus	0.00e+00	
25	515	33.8	1484	79	HPU34169	Human papillomavirus	0.00e+00	
26	515	33.8	1484	15	HP34184	Human papillomavirus	0.00e+00	
27	514	33.7	1484	79	HPU34187	Human papillomavirus	0.00e+00	
28	514	33.7	1484	79	HPU34190	Human papillomavirus	0.00e+00	
29	514	33.7	1484	79	HPU34183	Human papillomavirus	0.00e+00	
30	514	33.7	1484	79	HPU34177	Human papillomavirus	0.00e+00	
31	514	33.7	1484	79	HPU34186	Human papillomavirus	0.00e+00	
32	514	33.7	1484	15	HP34186	Human papillomavirus	0.00e+00	
33	514	33.7	1484	79	HPU34166	Human papillomavirus	0.00e+00	
34	514	33.7	1484	79	HPU34175	Human papillomavirus	0.00e+00	
35	514	33.7	1484	15	HP34165	Human papillomavirus	0.00e+00	
36	514	33.7	1484	15	HP34177	Human papillomavirus	0.00e+00	
37	514	33.7	1484	15	HP34187	Human papillomavirus	0.00e+00	
38	514	33.7	1484	15	HP34175	Human papillomavirus	0.00e+00	
39	514	33.7	1484	15	HP34183	Human papillomavirus	0.00e+00	
40	514	33.7	1484	15	HP34190	Human papillomavirus	0.00e+00	
41	514	33.7	1484	79	HPU34165	Human papillomavirus	0.00e+00	
42	514	33.7	1484	15	HP34166	Human papillomavirus	0.00e+00	
43	513	33.7	1484	79	HPU34193	Human papillomavirus	0.00e+00	
44	513	33.7	1484	79	HPU34179	Human papillomavirus	0.00e+00	
45	513	33.7	1484	79	HPU34192	Human papillomavirus	0.00e+00	

## ALIGNMENTS

RESULT	1	PAPHPV18	7857 bp	DNA	VRL	12-SEP-1993
LOCUS		Human papillomavirus type 18 genome.				
DEFINITION		X05015				
ACCESSION		60975				
NID		genome; overlapping genes; unidentified reading frame.				
KEYWORDS		Human papillomavirus.				
SOURCE		Human papillomavirus				
ORGANISM		Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus				
REFERENCE		1 (bases 1 to 7857)				
AUTHORS		Cole, S.T. and Danos, O.				
TITLE		Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products				
JOURNAL		J. Mol. Biol. 193 (4), 599-608 (1987)				
MEDLINE		8728382				
COMMENT		Data kindly reviewed (14-AUG-1987) by Danos O.				
FEATURES		Location/Qualifiers				
		source				
		1..7857				
		organism="Human papillomavirus"				
		misc_feature				
		87				
		/note="E6 ORF start site"				
		105..581				
		/note="E6 (AA 1-158)"				
		/codon_start=1				
		/db_xref="PID:g60976"				

/translation="MARFEDTPRRPKLPDLCTELNLSLQDIEITVCYKTVLEITEV  
FEAFKDLFVVYRDSIPHAACHKCIDFYSRIRLHSHYSDVYGDITLEKLTNTGLYNLL  
IRCLRCQKPLNPAEKRLHNAKRRFHNIAGHYRGQCHSCCNRRARQERLORRRETQV"  
509  
misc\_feature  
CDS  
/note="E7 ORF start site"  
590..907  
/note="E7 (AA 1-105)"  
/codon\_start=1  
/db\_xref="PID:g60977"  
/translation="MHGPKATLQDILVHLEPONEIPVDLLCHBQLSDSEENDEIDGV  
SQHLLPARRAEPORHTMLCNCCKCEARIELVLESSADDLRAFOQLFNLTSFVPCWCA  
908  
misc\_feature  
CDS  
/note="E1 ORF start site"  
914..2887  
/note="E1 (AA 1-657)"  
/codon\_start=1  
/db\_xref="PID:g60978"  
/translation="WADPEGIDEGTCGNGWYVQAIYDKTGDISSDDENADTGG  
SDWYDIDTGGTFCBOALEFAKALFAHGEVHNDQVLYLKRKFAAGSTENSLPGER  
LEVDTLSPLQELISLNGOKKAKRRLLTISDSGIGGSEVERATQIQTINGSHGNVC  
SGSTSAIDNGTEGNSVDGSDNSNIENVPQCTIAQLKOLLKYNKQGMJAVP  
KDTYGSFIDLVRNFKSDKTTCDWTAIFGVNPTIAEGKTLIQPILYAHQCLDC  
KWGVLLALRYKQKSLYAKGLSTLLHVPETCMLIQPPKURSSVAALYMYRTGIS  
NISEVMDPEWIKLITIIQIGDSDSFDLSEWQAFDNLDELSDMAFAYALLADS  
NSNAAPLNSCOAQAKDKATCMCKHYRRAQRQNMNSQWIRPCSKIDDEGDMRPV  
QFLRYQOIEFTFLGALKSLFKPKKNCVFCGPANTGSKSYFGMSFIHFIDGAVISF  
VNSTSHFLEP.LTDTKVAMLDATTCWTYFTYMRNALDGNPISIDRKHKLPIQLKC  
PPIILLTNIHPAKDRNPYLESRITVFEPNAPFPDKNGNPVYEINDKNKCKFFERTW  
SRLLDHEEBEADTEGNPFGTKLRAGQNHRL"  
2796  
/note="E2 ORF start site"  
2817..3914  
/note="E2 (AA 1-365)"  
/codon\_start=1  
/db\_xref="PID:g60979"  
/translation="NQTPKETLSRLSCVQDKIIDIHENDSKDISQIQYWOLIRWEN  
AIFFAAREHIGIOTLHNOVVPAYNISKSAKHAIELOMALQAGLSRYKEDTFLQDTC  
EELWNPTEGPKFKGQVQVYFDGNKDNCTIYVAMDSVYMDAGTWKDTKCVSHR  
GLYVREGNTYIFEFSECKYNGTWEVHFNNVIDCNSMCTSDTYSATQVNL  
KQLQHPYSPYSTVSGTAKTYGOTSAATPGHCGLAERKHCGVNPVLLGATPGNN  
KRLKSGNNTPIIHLKGDRLSKCLRYLRKSHDYRDISSTWHWTGAGNEKTKILT  
VYHSSTORTKFLNTVAIPDSVQILVGYMTM"  
3405..3409  
/note="E4 ORF start site"  
3418..3684  
/note="E4 (AA 1-88)"  
/codon\_start=1  
/db\_xref="PID:g60980"  
/translation="WLCZAVPTVTRYPLLSLNSYSTPPHRIAPCPWPAPQPTARRR  
LLHDLTVDSRSSHVSIDLSHFSVOLHLQATFKDONSVVYLRLL"  
3915  
misc\_feature  
CDS  
/note="E5 ORF start site"  
3936..4157  
/note="E5 (AA 1-73)"  
/codon\_start=1  
/db\_xref="PID:g60981"  
/translation="MLSLIFLFCPCVCMVYVCHVPLLPVCMCMYANWLVIYVIVIT  
SPATATVIVFCFLPLPMLLHAILSLQ"  
4166  
misc\_feature  
CDS  
/note="L2 ORF start site"  
4244..5632  
/note="L2 (AA 1-462)"  
/codon\_start=1  
/db\_xref="PID:g60982"  
/translation="MVSRRARRKRAVSIDLKTKCKSGCTPPDPVVKVEGTLLADKI  
LQWSSLIGLGLGIGTSGGRTGYPLGGKSNVYVDVGRTPPVVIEPVGPVDPDS  
IVTLIEDSSVYVSGAPRFTFTGSGEDITSAGTTPAVLDITPSSSVSISNTFNP  
AFSDPSLIEVPGTGEVAGNVPFTSGTHGYEEIPLQTFASSGTGEETISSTPLPTV  
RRVAGPRLYSRAIQVSVANPEFLTPSSLIYDNPAPFVDITLAFDRPSVDPDSF  
MDIILRHRLPALSRRGTVFRSGQRATMFTKSGTQIGARVHYHDIISPASPYIE  
LQPLVATENDLDFDIYADDNDPAVPVPSRSTSPFAFKYSPITSSASSYSNVYPLT  
SSMDVPVYTGPDITLPTSTSWPIVSPASTAQYIGIHGTHYLLMPLFYFIPKRRKR

VPYFFADGFAA"  
5418  
misc\_feature  
CDS  
/note="L1 ORF start site"  
5430..7136  
/note="L1 (AA 1-568)"  
/codon\_start=1  
/db\_xref="PID:g60983"  
/translation="NCLVTRVLIHLXPLVHPRPLHSLILVMMVHIIICGH  
YIILFNRNVNRPPIFLQMALWRPDSNTVLPSPSVARVNTDDYVTPSTISFYHAGSSR  
LLTVGVEIRVAGGKNQDIPKVSAYQVRVFRVOLPDNKFGLPDTISYNPRTQLRV  
WACGVEIRGROPAGVLSGHPFYNKLDPTESHAAATSNVSDVDNVSVDYKOTQLV  
ILCGAPAEHAKGTACKRPLSQGDCPPELKNVLEDDMDVDTGCGAMDPSLTQD  
TKCEVDLDCQSTCKYDPYLOMSADPYGSMFECLRRQLFARHFNRRACMGDTVPQ  
SLYIKGTGMPASPGSCVYSPSPSGSIVTSDSOLFKNKPYHLKHAQGNNGVCMHQLFV  
TVVDTPTSNLTIACASTQSPVGGQZDATERKQXSRHVEYDLOLIFOLCITITADVM  
SVTHSNSSITLEDNMFQPPPTPSLVDYRVQSVAITCQKDAAPAKENKDPDKLKF  
WVNDLKEKSLDLDDQYPLGRKFLVQAGLRKKTIGPKRSAPSATTSKPAKRVRYRA  
RK"  
BASE COUNT 2365 a 1497 c 1680 g 2315 t  
ORIGIN  
Query Match 97.48; Score 1484; DB 80; Length 7857;  
Best Local Similarity 98.7%; Pred. No. 0.00e+00;  
Matches 1504; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
Db 5613 atggccttbtggcgccctagtgacaataccgttatcttccacccctcttctgtggcaaga 5672  
QY 1 ATGGCTTTGTGGCGCCCTAGTGACAATACCGTATACCTTCCACCTCCTTCTGTGGCAAGA 60  
Db 5673 gttgtaataaccgtagtatgtgactcccaacacatatttatcatgctggcagctct 5732  
QY 61 GTTGTAATACTAGTATGTGACTCGCACACACATATTTATCATGTGGCAGCTCT 120  
Db 5733 agattataactgttggtaataccatatttagggttccctgcaggtggtggcaataaagcag 5792  
QY 121 AGATTATTAACTGTGTGTAATCCATATTTTAGGGTTCTTCGAGGTGGTGCATAAGCAG 180  
Db 5793 gatctcctaaggcttctgcataccaataagatatttaggtggtgcagttacctgaccca 5852  
QY 181 GATATCTTAAGGTTCTTCGATACCAATATAGAGTATTTTCGGGTGCGAGTTACTGACCCA 240  
Db 5853 aataaataattggtttaccctgatactagattattataactcctgaacacaaagtttagtggg 5912  
QY 241 AATAAATTTGGTTTACCTTGATAATAGTATTTATTAATCTGAAACACACAGTTTAGTGTGG 300  
Db 5913 gctctgtcgtggagtggaaattggcgtggtcagcctttagtggttggccttagtgggcat 5972  
QY 301 GCCTGTGCTGGAGTGGAAATTTGGCGGTGCTGAGCCTTTTAGTGTGTGGCCTTAGTGGCAT 360  
Db 5973 ccattttataataaattagatgacactgaagttcccatgccgcacgctctaattttct 6032  
QY 361 CCATTTTATAATAATTAGATGACACTGAAAGTTCCCATGCCGTACGCTAATGTTCT 420  
Db 6033 gaggacgttagggacaattgtctgtgattataaagcacacacagttatgtattttgggc 6092  
QY 421 GAGGACGTTAGGGACAATGTCTGTAGATTATNAGCAGACACAGTTATGTATTTTGGGC 480  
Db 6093 tgtgccccctgctattgggggaacactgggctaaaggcaactgcttgaataatcgctcttta 6152  
QY 481 TGTCCCCCTGCTATTGGGGAACACTGGGCTAAAGCACTGCTTGTAAATCGGCTCCTT 540  
Db 6153 tcacaggggcagattggccccctttagaaccttaaaacacagtttgggaagatgggtgatag 6212  
QY 541 TCACAGGCGCATGTGCCCCCTTTAGAACCTTAAACACACAGTTTGGGAAGATGTTGATG 600  
Db 6213 gtagatactgggatatggtgcactggacttttagtacattgcaagatactaaatgtgaggtta 6272  
QY 601 GTAGATTAGGATATGGTGCCATGGACTTTTAGTACATTGCAAGATATAAATGTAGGTA 660  
Db 6273 ccattggatattgtcagctctcattgttgaatatccctgattatttacaatgtctcgagat 6332  
QY 661 CCATTGGATATTGTGCACTATTGTGAAATATCTGTGATATTATACAAATGCTCGAGAT 720



CDS 5330..7149  
 /gene="li"  
 /codon\_start=1  
 /product="late protein"  
 /db\_xref="PID:g937029"  
 /translation="MAHNIYGHIIIFLNKVVNFFIFLQMLWRPSDSTVILPPSV  
 ARVSDYVSRSLITGVNPFYRVFVGNAGNQAQVPRVSAIQVRFV  
 ALPDNKEQDLEININPQRLVACVMEIGRGPLGLSGHPFFYNKLDPTESAH  
 AATVITQVDRNVDYQTLCLGCVPAIGHWAKGLGKLAQQLQDPLBLK  
 NIIIEEDGVNDYGVANDSTLDQKCEVPLDQICKYPLQMSADYVIGDSMEFC  
 LAREQLFARHFNWAGVMGDVPTDLYIKGTSAMRETPGSCVSPSPSIIISDSQ  
 LKPKYVLRHAKNGICWHQNLQVIVVDVTRSTNLCLCASTQNPVSIYDTPKEQ  
 YSRHVEYDLQFIFQCLITLAEVMSIENSMNSLENWFGVPPPTSLVDYTFE  
 VQSVAVTCQKDTTPPEKQDPYOKLFWVDLKEFSSDLDDPLGRKFLVQAGLRRP  
 TIGPKRPAAGSTFASRPAKRVIRSKR"  
 BASE COUNT 2409 a 1462 c 1652 g 2335 t  
 ORIGIN

Query Match 60.5%; Score 922; DB 79; Length 7858;  
 Best Local Similarity 82.3%; Pred. No. 0.00e+00;  
 Matches 1231; Conservative 0; Mismatches 255; Indels 9; Gaps 4;

Db 5608 atggccttggcggcctagtcacagtcaggtatcttccacacattctgtggcaga 5667  
 QY 1 ATGGCTTGTGGCGCTAGTCACAAACGGTATACCTTCCACCTCCTCTGTGGCAGA 60

Db 5668 gtgtgcagcactgatgatattgtctgcgcagaagcatatttttcatcagcgacttcc 5727  
 QY 61 GTTGTAAATGATGATGTGACTGCGCAAGCATATTTATCATGTGGCAGCTCT 120

Db 5728 cgattctaaccttgaggcaatccatatttttaggggtgtacctaaatgggcaggtataaa 5787  
 QY 121 AGATTATTAATCTGTGGTAAATCCATATTTTAGGGGT---CCTGCAGGTGGCAATAAG 177

Db 5788 cagggtgttctaaggtataccgcatacatcagtcagtgagggttttagtagtagtettaccogat 5847  
 QY 178 CAGGATATTCCTAAGGTCTTCATACCAATATAGAGATTTTCGGGTGCAGTTACCTGAC 237

Db 5848 cctaataaatttggaatacctgattctactatataaactgcgaacacacacttttggtt 5907  
 QY 238 CCAAAATAAATTGGTTTACCTGATAATAGTATTTATAATCCTGAACACACACGTTTAGTG 297

Db 5908 tgggcattgtcaggtatggaaattggtctggcgcagcccttaggtattggcctcaatgggc 5967  
 QY 298 TGGGCTGTGCTGGAGTGAAATTTGGCGGTGGTCAGCCCTTAGGTGTTGGCCTTAGTGGG 357

Db 5968 catccatttataataaatttgatgatacagaaagtgctcatgcagctcagctgttatt 6027  
 QY 358 CATCCATTTTATAAATTAGATGACACTGAAAGTTCCCATGCCGCTACGTCTAATGTT 417

Db 6028 acgcaggatgtcagggaataatgtccagttgattataaagcaaacacagctgtatttta 6087  
 QY 418 TCTGAGGAGCGTTAGGCACAATGTGCTGTAGATATTAAGCAGACACAGTATGTATTTTG 477

Db 6088 ggttgtacatcgtattgttagcaactggccgcgaaggcgacactttgtaaacctgcacaa 6147  
 QY 478 GCGTGTCCCCCTGCTATTTGGGACACTGGGCTAAGGACACTGCTGTGTAATGCCGCTCT 537

Db 6148 ttgcacactggtgactgctcctccttggaaacttaaaacacacattatcgaggtggtat 6207  
 QY 538 TTATCACAGGCGGATTTGCCCCCTTTAGAACTTAAGACACAGATTTTGAAGATGTGTGAT 597

Db 6208 atggtggatacaggtattggggcaatggatttagtcattcattcaggatatacaaatgctcga 6267  
 QY 598 ATGGTAGATACGTGATATGTCCTAGGACCTTTAGTACATATGCAAGACTACTAATGTGAG 657

Db 6268 gttccattagacatttgtcaaccatcctctgtaaatatccagattatttgcacatgcttgt 6327  
 QY 658 GTACCATTTGGATATTTGTGAGTCTATTTGTAAATATCCTGATTTATTAATAATGCTGCA 717

Db 6328 gatccctatggggattctatgttttttggcctacgcgcgtgaacaaactgtttgcaagacat 6387  
 QY 718 GATCCTTATGGGATTCATGTTTTTTTGTCTACGAGCTGAGCAGCTTTTGTGTAGGCAT 777

Db 6388 ttttggaaatgggcaggtgttatgggtgacacaggtacctaoggaacctatataaaagc 6447  
 QY 778 TTTTGGAAATAGGGCAGGTACTATGGGTGACACTGTGCCCAATGCTTATATATAAAGC 837

Db 6448 actagcgctaataatgcgtgaaacccccggcagttgtgttatcccttctccacagtgc 6507  
 QY 838 AC-AG- - - - -TATGGGTGCTTACCTCGGCAGCTGTGTATCTCCCTCTCCAAGTGC 891

Db 6508 tctattactctctgattctcaattatttaaagaagccatttggtacataaagggccag 6567  
 QY 892 TCTATTGTACTCTGACTCCCGAGTGTGTTAAATAAACCATATATGTTGTTACATAAAGGCACAG 951

Db 6568 ggcatacaaatggattttgttggcacaataacagttgttggtaactgtagtggacactacc 6627  
 QY 952 GGTATACAAATGATATCTGCTGGCATAATCAATATTATTGTACTGTGTGATACCACT 1011

Db 6628 ccagctactaatttaacattatgtgcctctcacacaaaacccctgtgcceagtcacattgac 6687  
 QY 1012 CGTAGTACCAATTTAAACAATATGCTTCTACACAGCTCTCCGTACCTGGCAATATGAT 1071

Db 6688 cctactaagtttaagcagctatagtcagacatgtgsggaataatgattttcacagttattttt 6747  
 QY 1072 GCTACCAAAATTAAGCAGTATAGCAGACATGTTGAAGAATATGATTTGCAGTTATTTTT 1131

Db 6748 cagttgtgcactattactttaactgcagaggttatgtcctatataccatagtagtaagtagt 6807  
 QY 1132 CAGTTATGTACTATTACTTTAACTGCAGATGTTATGTCCTATTTCAATAGTATGAATAGC 1191

Db 6808 agtatcagaaaattggaaatttgggtgcctccacacccctcactacaagtttggtagt 6867  
 QY 1192 AGTATTTTAGAGATTTGGAACTTTGGTGTTCCTCCCTCCCTCCCTACCTAGTGTGGTGGAT 1251

Db 6868 acatatcgttttgcacacagttgctgttaccgtgcacaaaggaatactcacacctccagaa 6927  
 QY 1252 ACATATCCTTTTGTACAAATCTGTGCTATTACCTGTCAAAAGGATGCTGCACCAGCTGAA 1311

Db 6928 asgcaggatcccatgataataataaagtttggactgttgacctaaagaaagaaatcttc 6987  
 QY 1312 AATAAGGATCCCTATGATGAATGAATGAATTTGGAATGTGGATTTAAAGAAAGATTTTCT 1371

Db 6988 tccgatttggatcaaatcccttgggtgcgaagtttttagtttcagcgtgggttgcgtcgt 7047  
 QY 1372 TTGGACTTAGATCAATATCCCTTGGAGCTAAATTTTGGTTCAGCGCTGGATTGCGTCGC 1431

Db 7048 aggcctacataagcactgtaagtcctgcctgcctccacagctcactacgtcactacgtacta 7102  
 QY 1432 AAGCCACCATAGGCCCTCGTAAACGTTCTGCTCCATCTGCCACTACGCTCTCTTA 1486

RESULT 3  
 LOCUS HPV59YG 7896 bp DNA VRL 10-OCT-1994  
 DEFINITION Human papilloma virus type 59, complete viral genome.  
 ACCESSION X77858  
 NID 5557236  
 KEYWORDS complete genome.  
 SOURCE Human papillomavirus type 59.  
 ORGANISM Human papillomavirus type 59.  
 Viruses; dsDNA viruses, no RNA stage; Papovaviridae;  
 Papillomavirus.  
 1 (bases 1 to 7896)  
 Rho, J., Roy-Burman, A., Kim, H., de Villiers, E.M., Matsukura, T. and  
 Choe, J.  
 Nucleotide sequence and phylogenetic classification of human  
 papillomavirus type 59  
 JOURNAL Virology 203 (1), 158-161 (1994)  
 MEDLINE 9430329  
 REFERENCE 2 (bases 1 to 7896)  
 AUTHORS Choe, J.  
 TITLE Direct Submission  
 Submitted (25-FEB-1994) to the EMBL/GenBank/JDBJ databases. J.  
 Choe, C/O Ha Jo Delli, DKFZ - Abt. ATV, Im Neuenheimer Feld 506,  
 69120 Heidelberg, FRG

Location/Qualifiers	
1..7896	/db_xref="PID:g557244"
55..537	/db_xref="PID:g557238"
872..2806	/db_xref="PID:g557237"
542..865	/db_xref="PID:g557243"
5606..7132	/db_xref="PID:g557241"
6445	/db_xref="PID:g557240"
6385	/db_xref="PID:g557239"
6325	/db_xref="PID:g557238"
6265	/db_xref="PID:g557237"
6205	/db_xref="PID:g557243"
600	/db_xref="PID:g557241"
6085	/db_xref="PID:g557240"
420	/db_xref="PID:g557239"
360	/db_xref="PID:g557238"
300	/db_xref="PID:g557237"
5905	/db_xref="PID:g557243"
5965	/db_xref="PID:g557241"
6025	/db_xref="PID:g557240"
420	/db_xref="PID:g557239"
6085	/db_xref="PID:g557238"
480	/db_xref="PID:g557237"
6145	/db_xref="PID:g557243"
540	/db_xref="PID:g557241"
6205	/db_xref="PID:g557240"
600	/db_xref="PID:g557239"
6265	/db_xref="PID:g557238"
660	/db_xref="PID:g557237"
6325	/db_xref="PID:g557243"
720	/db_xref="PID:g557241"
6385	/db_xref="PID:g557240"
780	/db_xref="PID:g557239"
6445	/db_xref="PID:g557238"
6445	/db_xref="PID:g557237"

Location/Qualifiers	
1..7896	/db_xref="PID:g557244"
55..537	/db_xref="PID:g557238"
872..2806	/db_xref="PID:g557237"
542..865	/db_xref="PID:g557243"
5606..7132	/db_xref="PID:g557241"
6445	/db_xref="PID:g557240"
6385	/db_xref="PID:g557239"
6325	/db_xref="PID:g557238"
6265	/db_xref="PID:g557237"
6205	/db_xref="PID:g557243"
600	/db_xref="PID:g557241"
6085	/db_xref="PID:g557240"
420	/db_xref="PID:g557239"
360	/db_xref="PID:g557238"
300	/db_xref="PID:g557237"
5905	/db_xref="PID:g557243"
5965	/db_xref="PID:g557241"
6025	/db_xref="PID:g557240"
420	/db_xref="PID:g557239"
6085	/db_xref="PID:g557238"
480	/db_xref="PID:g557237"
6145	/db_xref="PID:g557243"
540	/db_xref="PID:g557241"
6205	/db_xref="PID:g557240"
600	/db_xref="PID:g557239"
6265	/db_xref="PID:g557238"
660	/db_xref="PID:g557237"
6325	/db_xref="PID:g557243"
720	/db_xref="PID:g557241"
6385	/db_xref="PID:g557240"
780	/db_xref="PID:g557239"
6445	/db_xref="PID:g557238"
6445	/db_xref="PID:g557237"

```
QY 781 TGGATAGGCGAGTACTATGGGTGACACTGTGCCTCAATCTTATATTAAGGCACA 840
Db 6446 gacatcgtgccaaccaggagctattattatattcccttcccagggtgctgtggtt 6505
QY 841 GGTATCGGTGCTTACCTGCGCAGCTGTGTATTCTCCCTCCTCAAGTGGCTATGTT 900
Db 6506 acttcgattcacaattattataaaccatattgtctgcacaggctcagggtttaaac 6565
QY 901 ACCTCTGACTCCAGTGTGTTAATAAACCATATTGGTTACATAAGGCACAGGTCATAAC 960
Db 6566 aatggtatattgtgcacacatcaattgttttaacagcttagatactactcagcagcacc 6625
QY 961 AATGGTATCTGCTGGCATAATCAATTATTGTTACTGTGGTAGTACCACTCGTAGTACC 1020
Db 6626 aatcttttctgtgtctgtctactactcttcttcttcttcttcttcttcttcttctt 6685
QY 1021 AATTTAAACAATATGTCCTTACACAGTCTCTCTGTACCTGGGCANATATGATCTACDMAA 1080
Db 6586 tttaagaatatgcagacatcgttgggggaatttggatttggagtttatttccaactgtgt 6745
QY 1081 TTTAAGCAGTATAGCAGACATGTGTGAAGAAATATGATTTGCAGTTTATTTTTCAGTTATGT 1140
Db 6746 aaataaacttaactacagaggtaagtcatcacatcattoataatgaataccactattttg 6805
QY 1141 ACTATTACATTTAACTGCAAGATGTATGTCTATATTCATATGATGAATAGCAGTATTTTA 1200
Db 6806 gaggattggaatttgggtgttacacacccctcctactgttagtttagtgacacataccgt 6865
QY 1201 GAGGATTGGAACCTTTGGTGTTCCTCCCTCCCGCAACTACTAGTTTGGTGTATACATATCGT 1260
Db 6866 ttgttcaactcgtcgtcgtgaactttgtcaaaaggacacccgcagcttaaacagagac 6925
QY 1261 TTTGTACAATCTGTGCTATTAGCTGTCAAAAGGATGTCGCACACAGCTGAAAAATAAGGAT 1320
Db 6926 ccttatgacaaactaaagtgttggcctgtagatcttaaggaaaggttttctgcagattt 6985
QY 1321 CCCTATGATAAAGTTAAAGTTTGGAAATGTTGATTTAAAGGAAAAAGTTTCTTTGGACTTA 1380
Db 6986 gatcagttcccttggagcgtcaaatcttctgcaattagagctagagctagagcctaagccact 7045
QY 1381 GATCAATATCCCTTGGACGTAAATTTTGGTTCAGGCTGGATGGCTGCACAGGCCACACC 1440
Db 7046 ataggccacgcaacgtgcagcgc 7071
QY 1441 ATAGGCCCTCGTAAACGTTCTGCTCC 1466
```

```
4
RESULT
LOCUS      HPV11      1855 bp      DNA      VRL      16-FEB-1996
DEFINITION Human Papilloma Virus L1 gene for major capsid protein.
ACCESSION X67161
NID        g1197494
KEYWORDS   major capsid protein.
SOURCE     Human papillomavirus type 68.
ORGANISM   Human papillomavirus type 68.
            Viruses; dsDNA viruses, no RNA stage; Papovaviridae;
            Papillomavirus.
REFERENCE  1 (bases 1 to 1855)
AUTHORS   Longuet,M.
TITLE     Direct Submission
JOURNAL   Submitted (03-JUL-1992) M. Longuet, Institut Pasteur, Laboratoire
          des Papillomavirus, 25 Rue du Dr Roux, 75024 Paris Cedex 15, FRANCE
REMARK    Revised by author 05-DEC-95
REFERENCE  2 (bases 1 to 1855)
AUTHORS   Longuet,M., Beaudenon,S. and Orth,G.
TITLE     Two Novel Genital Human Papillomavirus (HPV) Types, HPV68 and
          HPV70, Related to the Potentially Oncogenic HPV39
JOURNAL   J. Clin. Microbiol. 34, 738-744 (1996)
FEATURES   Location/Qualifiers
            1..1855
            /organism="Human papillomavirus type 68"
            /cell_line="intraepithelial neoplasia"
            194..1711
            CD5
```

```
BASE COUNT      501 a      356 c      398 g      600 t
ORIGIN
Query Match      48.6%; Score 740; DB 79; Length 1855;
Best Local Similarity 75.0%; Pred. No. 0.00e+00;
Matches 1111; Conservative 0; Mismatches 371; Indels 0; Gaps 0;
Db 194 atggcattgtggcgagctagcagacaacatggtgtatttgcctccccctccatggtgcgaag 253
QY 1 ATGGCTTTGTGGGGCCCTAGTAGACATACCGTATACCTTCCACCTCCTCTCTGTGGCAAGA 60
Db 254 gttgcatacagatgattatgtgacacgcactggcattgtattactatgctggtacatct 313
QY 61 GTTGTAAATACATGATGATTGTGACTCGCACAAAGCATATTTATCATGTGCGCAGCTCT 120
Db 314 aggttatttaactgtgagccatcccatatttaaggttccctatgctctgggggcccgaagcag 373
QY 121 AGATTATTAACTGTGTGGTAATCCCATATTTTAGGGTTCTCGCAGGTGTGGCAATAGCAG 180
Db 374 ggcattcctaaaggtgtcgcatacacaacagagtggttttaggttaaccttacctgacct 433
QY 181 GATATTCCTAAGGTTCTGTCATACCAATATAGAGTATTTCGGGTGTCAGTTACCTGACCCA 240
Db 434 aataaatttagtctcctgagctcacattatataatccagatacacagcgcatggtatgg 493
QY 241 AATAAATTTGGTTTACCCTGATAAATAGTATTATTAATCTCTGAAACACAACTTTAGTGTGG 300
Db 494 gccctgttgggttgaaataggtagggggcagccattggcgttggccttagtggcat 553
QY 301 GCCGTGCTGGAGTGGAAATGGCCGTGTGTCAGCCTTTTAGTGTGTGGCCTTAGTGGCAT 360
Db 554 ccactatataagctggtgagtagactgatactgaaaaattccccgttttccctctaataaaaatcct 613
QY 361 CCATTTTATAAATTAGATGACACGTGAAGTCCCATGCCGCTACGTCTAATGTTCT 420
Db 614 aaacacagtaggacaacatgttgcagtgactgtatacaacaacacagctgtgtattataggc 673
QY 421 GAGGACCTTAGGGACANATGTGTCGTAGATTATTAAGCAGACACAGATTATGATTTTGGGC 480
Db 674 tgttctcgtctattggcgagcactggggccaaagtaaatcttgaagcctaccacatgta 733
QY 481 TGTGCCCTCTATTGGGGAACACTGGCTTAAAGGCATGTCTTGTAAATCGCTCCCTTTA 540
Db 734 caacaaggagcactgtcccccatttggaaattggtaaatctcctctatttggaggtggcgatatg 793
QY 541 TCACAGGGCGATTGCCCCCTTTAGAACTTTAGAACACACAGTTTGGGAAGATGTTGATATG 600
Db 794 attgatacagatatggtgctactgagcatttgggtacattacaagaacacacagcaggtta 853
QY 601 STAGATACTGGATATGGTGCCATGGACTTTTAGTACATTGCAAGATACTAAATGTGAGGTA 660
Db 854 ccttggatatatgcaatcctggttgcaaatatcctcgactatttggcaaatgtctgcagat 913
QY 661 CCATTTGGATATTTGTCAGTCTATTGTAAATATCTCTGATTTATTTACAAATGTCTGCAGAT 720
Db 914 gtgattggagacagtagtgttttttggtttgcgttagggaacaggttatttgcaggcatttt 973
QY 721 CCTTATGGGATTCATGTTTGTGCTTACGACGTGACGACGCTTTTGTGCTAGGCAATTT 780
Db 974 tggaaatgggagcagtggttagggggaacactattccccactgacatgtattataaggcact 1033
```

```
/gene="L1"
/codon_start=1
/product="major capsid protein"
/db_xref="PID:g1197495"
/translaton="MALMRKSDNMYIYPPPSVAKVVTDDVYVTRTGMYYAGTSLRLT
VGHFVPMGSGRGKQIPKVSADNYVFRVLPDPNKFSPSTLXNPOTQRMVAC
VGVIGGQPLGVLGSHPLNRLDDTENSFFSNKNPKDRDNVAVDCKTOLCIIIG
CVAIGHMAKGSCKPTNVQGGCPPLLELVNTPIEDGMDTGYGMDRGTOLCTKS
EVLIDICQSVKYPDYLONSADVYGDMSFFCLREQLFAHFNHNRGNGMVDLTPTMY
IKGTDIRETPSSVYAPSPSGSWSSDQLFNKPYWLHKAQHNHNGICWHNOLFVW
DITRSTNFTLSTTDTSTPAVYDTSNKFKEVRYRVEYDLQRTFOLCITLSTDMYSYI
HTMNPALDDMNEGVAPPPSASLVDTYRYLQSAATTCOKDAPAPVKPDYDGLNFWNV
DLKEKFSSELDQPLGRKFLQAGVRRRPTIGPRKRTATATTSTSKRKRKRVSK"
```





IVOLVEBSVSSGTPIPTFTSGSGFETISSATTTPAVDITDPASGVSQISTSYNPN  
ADAPSLIEVTEGEVSNVETITSTGTHGYEINQWFAHSGTSTETSPSTPFGV  
SRVAPGYSRAYHQVNNEDFVTPSEFTEVPFAPGSGTSTETPADATPDPG  
LDIVLRHLPATSRGVRVAGNEMKHTTRGTQGAQVHYHIDINSTATEDIED  
QPLIERSDGLJLIDYAADADINLAKHTTSGTSGRPSHLSPGTSVSTYKNT  
IPTTSDWIDIPVLTPTASPNLPVPPPTSIDTTVAITAGGNSVYLLPPLLYELK  
KRYEIVTEPTEGVAV\*

CDS

5590..7104 /gene="I.1"

```

/gene= LI
/note="Method: conceptual translation supplied by author.;
putative LI"

```

putative LI  
/codon\_start=1  
/db\_xref="pid:g1160346"

```
/usr_xref=FID:g1160346
/translation="MALWRSSDNTVLPSPSAKVNTDDYVTRTGIIYYAGSSRLLT
VGHPEYKVPVNGGCKQFIKVSAYQYVPEPVSIPDNKEGIQDDPSIYNBPTQRIYVAC
```

VGHFIFKVPNGGKQKLEIPKVSIAIKVFKVSTLFDNRFGLFDFSLINFDIQLRLVWA  
 IGVEIGRGQPLGVSGHPLYNRLLDDTENSHESSAVSTQDSRDNVSDYKQTKLCIIG  
 CVPAMGEHWAKKCKASTOOGNCBBI,FLVNTAIFEDGMDITGVGAMVMTI,ETKSEV  
 VGHFIFKVPNGGKQKLEIPKVSIAIKVFKVSTLFDNRFGLFDFSLINFDIQLRLVWA

CVPAMGEHWAKGRACKRSTQGGDCFFTELNVNIAEEDMDTJGAMDFRTLQETKSE  
 PLDQCQSWCKYDYLQMSADYVGDSSMFLKQELFARHWNRGGMVGTIPSELYIK  
 STDIPEEBGCTHVSBSBSCSMVYSSSOLFNKZBPYWLHKAQGNNGICENOLFETVVDTE

TRSTNETLSACTAIPAVSYPTKFEYEEYDQFIFQLCTITLTADVMAIYHT  
GDSIKRPGTHVTSFSPSGSMVSSDSQLFNKPIWLHKAQGHNNQICWHNQLEITVAVDT  
TSTNNTJNWNAGVTBDBSASIVDTYRVLQSAACOKNADTPEKKKBDLKEWNTDL  
MNPAT

KEKSTELDQFPLGRKFLQVGRRRPTIGPKRPPASAKSSSSASKHKRVRVK"  
1495 C 1693 C 2292 +

1499 C 1093 g 2292 C

47.58; Score 724; DB 79; Length 7905;  
75.58; 5253; 0.000+00.

ility 75.6%; Pred. NO. 0.00e+00;  
 nservative 0; Mismatches 352; Indels 6; Gaps 4;

ggcggtcttagtgacaacacacggtgtattgcccccttctgtggcggaag 5649

GGGGGCCTAGTGACAAATACCGTATACCTTCCACCTCCTTCTGTGGCAAGA 60

ccagatgattatgtaacacgtacaggcatatattattatgctggaagctct 5709

CTGATGATTATGTGACTCGCACAAGCATATTTTATCATGCTGGCAGCTCT 120

ccagtagggcatccttattttaagggtacctgtaaatgggtggccgcaagcag 5769

CTGTTGGTAATCCATAATTTAGGGTTCTGCAGGTGGTGGCAATAAGCAG 180

Db 6244 ccactagatatttgccaaatccgtgtgtaaataatactctgattatttgcagatgtctgtctgat 6303

661 CCATTGGATATTGTCAGTCATTGTAATAATCCTGATTATTACAAATGCTGCAGAT 720

Db 6304 gtatatggggacagtatgtttttttgttttgcgaaggaaacagttgttttgcacgaccttt 6363

Qy 721 CCTATGGGATTCACATGTTTTTTTGGCTACGACGTGAGCAGCTTTTGTCTAGGCATTTT 780

Db 6364 tggaaatagaggtggcatggtggggcgacacaataaccttcagagtatatattaaaggcacg 6423

QY 781 TGAATAGGCAGGTACTATGGGTGACACTGTGCTCAATCCTTATATATTAAAGGCACA 840

Db 6424 gatatacgtgagcgtcctggtactaatgtatattcccccttcccccaagtggctctatggtc 6483

QY 841 GGTATGCGGCTTCACCTGGCAGCTGCTGTATCTCCCTCTCCAAGGGCTCATTTGT 900

Db 6484 tcttctgattccccagttgtttaataagccctattggttgataagggccaggacacaat 6543

QY 901 ACCTCTGACTCCCACTGTTTAATAAACCATAATGGTTACATAAGGCACAGGGTCATAAC 960

Db 6544 aatggcattgttggcataacgaagtgtttattactgtgttggaactacacgtagtact 6603

QY  
961 AATGGTATCTGCTGGCATAATCAATTATTGTTACTGTGGTATACCACTCGTAGTACC 1020

Db 6604 aatttacattgtcgcctgcaccgaaaggccatcctgctgtatatagccctacaag 6663

QY 1021 AATTTAACAAATATGTGCTTCTACACAGTCTCCTGTACCTGGSCAAATATGATGATACCAA 1080

RESULT	6		VRL	22-MAR-1994
LOCUS	PPHT39	7833 bp	DNA	
DEFINITION	Human papillomavirus ORFs.			
ACCESSION	M62849	M38185		
NID	G333245			
KEYWORDS	intraepithelial neoplasia; invasive carcinoma; papillomavirus.			
SOURCE	Human papillomavirus (type 39) DNA.			
ORGANISM	Human Papillomavirus type 39			
	Viridae; ds-DNA nonenveloped viruses; Papovaviridae.			
REFERENCE	1 (bases 1 to 7833)			
AUTHORS	Volpers,C. and Streeck,R.E.			
TITLE	Genome organization and nucleotide sequence of human papillomavirus type 39			
JOURNAL	Virology 181, 419-423 (1991)			
MEDLINE	91135017			
FEATURES	Location/Qualifiers			





```
QY 121 AGATTATTAACTCTTGGTAATCCATATTTTAGGGTTCTCGCAGGTGGTGGCAATAAGCAG 180
Db 5823 gacattccaaagtgctgcatacaataatagggtatttcgctgacattgcccgatctt 5882
QY 181 GATATCTCTAAGGTTCTGCATACCAATATAGAGTATTTCCGGTGCACTTACCTGACCA 240
Db 5883 aataaattcagatttccagatgcatacttataataaccagaacaacaacatttagtatg 5942
QY 241 AATAAATTTGGTTACTGTAATAGTATTTATTAATCCCTGAACACACACGTTTAGTGCG 300
Db 5943 gcttgtaggggtgaggtgggcaaggggccacccattgggtggttattatagtagaac 6002
QY 301 GCTGTGCTGAGTGGAAATGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
Db 6003 ccattataataagacagatgatactgaaactcsc---catttccatacaaccacaaat 6059
QY 361 CCATTTTATATAATATAGATGACACTGAAGTCCCATGCCCTGACGTCTAATGTTCT 420
Db 6060 aagcacagtgaggataatgtctgtggattataaacacagacacagtggtgctattataggc 6119
QY 421 GAGGACGTTAGGACAAATGTCTGTAGATTATAAGCACACACAGTTATGTATTTGGGC 480
Db 6120 tgtgttccctccattgggagacatggggglaaggggaaagggcagcccaacaaatgta 6179
QY 481 TGTGCGCCCTGCTATTGGGAACACTGGGCTAAAGGCACCTGCTTGTAAATCGCTCCCTTA 540
Db 6180 tctacgggggacgtctcctcttggaactagtaaacacccctattggagatgggtgatg 6239
QY 541 TCACAGGGCGATTGCCCCCTTTAGAACTTAAGAACACACAGTTTGGAAAGATGGTGATAG 600
Db 6240 attgatactgcatagagctagacttgggtgctattgcaggaacacaaagtgaggg 6299
QY 601 GFAGATACGAGATGGTCCCATGACCTTAGTACATTGCAAGATACATAATGTGAGGTA 660
Db 6300 cctttagctattgtcaactccttggactgtaaacctcctgattatttcaaatgtctcgagt 6359
QY 661 CCATTGGATATTTGTCAGTCTATTGTAAATATCCCTGATTAATTTACAAATGCTGCAGAT 720
Db 6360 tggtaggacagctatgtctctgttcctagtaggaacaacttttgcaaacatttt 6419
QY 721 CTTATGGGATTCATGTTTCTTTAGTACGACGAGCAGCTTTTGTCTAGGCATTTT 780
Db 6420 tggagctgggtggtatgggtggaacccattcctgcccacatttatattaaaggacaa 6479
QY 781 TGAATAGGCGAGTACTATGGGTGACACTGTGCTCATATCTTATATTTAAGGCACA 840
Db 6480 gatatactgcaaaccccggtggtctgtatactgcccctctccacagcgggtccatggta 6539
QY 841 GGTATGCGTGTCTACCTGGCAGCTGTGTGTATCTCCCTCTCCAAGTGGCTCTATTGTT 900
Db 6540 acctctgattcccgattatttaataagccttattggctacataaaggcccgagccacaac 6599
QY 901 ACCTCTGACTCCCAAGTTGTTTAAATAACCATATTGGTTACATAAGGCACAGGGTCAATAAC 960
Db 6600 aatggtatattgtggcataatcaattatttcttactgtgtgagacactacccttagtacc 6659
QY 961 AATGGTATCTCTGGCATAATCAATTAATTTGTTACTGTGTAGATACCACCTCGTAGTACC 1020
Db 6660 aactttacattatctacctataagagctcttccatacctctacataatgatccctcttaag 6719
QY 1021 AATTTAAACAATATGCTTCTACACAGTCTCCTGTACCTGGGCAATATATGCTACCAAA 1080
Db 6720 tttaaggataataccagcagctggagagtagattatacaaatttatatttcaactgtgt 6779
QY 1081 TTTAAGCAGTATAGCAGATGTTGAAGAAATATGATTTTGCAGTATTTTTCAGTTATGT 1140
Db 6780 actgtcaattaaacaactatgtattgtcttatattcacaactatgaattcctctatattg 6839
QY 1141 ACTATTACTTTAAGTCAGATGTTATGTCCTATATTTCATGATGATGAATAGCAGTATTTTA 1200
Db 6840 gacaatttgaatttctgtactcctccacacatctccagcttggtagacactacaga 6899
QY 1201 GAGGATTCGAACCTTGGTGTTCCTCCCTCCCGCCAACTACTAGTATTGGTGGATACATATCGT 1260
Db 6900 tacctacagctctgcagccattaccatgtcaaaagatgctccagccctgaaagaaagat 6959
QY 1261 TTTGTACAAATCTGTGCTATTACCTGTCAAAAGGATGCTGCACAGCTGAAAAATAAGGAT 1320
Db 6960 ccattagcaggtctaaagtttggatgttgaacttaagggaagaaagtttagtttgaactt 7019
QY 1321 CCTATGATAGATTAAAGTTTGGAAATGCTGAATTAAGGAAAAAGTTTCTTTGGACTTA 1380
Db 7020 gataatccctctgggagagtaaatattttgtgcagcccaagggtccgcagggcgccact 7079
QY 1381 GATCATATATCCCTTGGACGTAAATTTTGGTTGAGCTGGATTCGGTCGCAAGCCCAACC 1440
Db 7080 ataggctcccaaaaggcgtcgtcgtacccacttctcgtcctc 7123
QY 1441 ATAGGCCCTCTAAACGTTCTGCTCATCTGCCACTAGCTCTTC 1484
7
RESULT HUMHPVME18 6042 bp DNA PRI 11-OCT-1991
LOCUS Human cellular DNA/Human papillomavirus proviral DNA.
ACCESSION M73258
NID 9184383
KEYWORDS proviral gene.
SOURCE Human papillomavirus DNA; and Homo sapiens cervix DNA.
ORGANISM Human papillomavirus
REFERENCE 1 (bases 1 to 6042)
AUTHORS Reuter,S., Delius,H., Kahn,T., Hofmann,B., Zur Hausen,H. and
Schwarz,E.
TITLE Characterization of a novel human papillomavirus DNA in the
cervical carcinoma cell line ME180
JOURNAL J. Virol. 65, 5564-5568 (1991)
MEDLINE 91374616
FEATURES
source 1..6042 Location/Qualifiers
source 1..5993 /organism="Human papillomavirus"
source 1..5993 /organism="Human papillomavirus"
mat_peptide 221..1633 /sequenced_mol="DNA"
mat_peptide /standard_name="L2"
mat_peptide 1584..3131 /codon_start=1
mat_peptide /standard_name="L1"
mat_peptide /note="Open reading frame"
mat_peptide 3925..4419 /codon_start=1
mat_peptide /standard_name="E5"
mat_peptide 4337..4759 /codon_start=1
mat_peptide /standard_name="E7"
mat_peptide 4763..5116 /codon_start=1
mat_peptide /standard_name="E1a"
mat_peptide 4960..6039 /codon_start=1
mat_peptide /standard_name="E1b"
source 5994..6042 /organism="Homo sapiens"
source /cell_line="ME180"
source /sequenced_mol="DNA"
source /tissue_type="cervix"
BASE COUNT 1810 a 1155 c 1279 g 1798 t
ORIGIN
Query Match 47.1%; Score 718; DB 59; Length 6042;
Best Local Similarity 74.2%; Pred. NO. 0.00e+00;
Matches 1100; Conservative 0; Mismatches 382; Indels 0; Gaps 0;
Db 1617 atggcatttgccgctctagcgacaacatgggtgattgttgcctccccctcagtgggcgaag 1676
```

[illegible]

QY	1081	TTTAGCAGTATAGCAGACATGTTGAAGAATATGATTTCAGTATTATTTTTCAGTTATGT	1140
Db	2757	actaaacatgtccactgatgaatgacccatatacatatacatgaatccctgctatttg	2816
QY	1141	ACTATTACTTAACTGCGAGATGTTATGTCCTTATTTCATAGTATGATAGCAGATTTTA	1200
Db	2817	gatgattggaatttgggtgtgcccctccaccatctgtagtctttagatatacataccgc	2876
QY	1201	GAGGATTGGAACCTTGTGTGTTCCCGCCCGCAACTACTAGTCTTGGTGGATACATACGT	1260
Db	2877	tatctgcaatcagcagcaattacatgtcaaaaaagcgcctcgcactactataaaggat	2936
QY	1261	TTTGACAAATCTGTGCTATTACCTGTCAAAGATGCTGCACCAAGCTGAAATAAGGAT	1320
Db	2937	ccatatgatgctcaaaccttttggaaagttaattaaaggaagaagttagtcttgaactg	2996
QY	1321	CCCATGATATAGTTAAAGTTTGGAAATGGATTTAAAGGAAAAGTTTCTTTGGACTTA	1380
Db	2997	gaccagtttctcttaggacgcaaatcttctttagcagcaggtgctccgcgacgaccact	3056
QY	1381	GATCAATATCCCTTGGACGTAAATTTTGTTCAGGCTGGATTGCGTCGCAACCCACC	1440
Db	3057	ataggcccccgtaaacgccttcgacagcaactactgcact	3098
QY	1441	ATAGCCCTCGTAAACGTCTGCTCCATCTGCCACTAGTCT	1482

RESULT	8	HPV26	7855 bp	DNA	VRL	26-MAY-1994
LOCUS		Human papillomavirus type 26		genomic DNA		
DEFINITION		Human papillomavirus type 26				
ACCESSION		X74472				
NID		g396956				
KEYWORDS		E1 gene; E2 gene; E4 gene; E6 gene; E7 gene; early protein; L1 gene; L2 gene; late protein.				
SOURCE		Human papillomavirus type 26.				
ORGANISM		Human papillomavirus type 26				
REFERENCE		Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.				
AUTHORS		1 (bases 1 to 7855)				
TITLE		Deliuss, H. and Hofmann, B.				
JOURNAL		Primer-directed sequencing of human papillomavirus types				
REFERENCE		Unpublished				
AUTHORS		2 (bases 1 to 7855)				
TITLE		Direct Submission				
JOURNAL		Submitted (06-AUG-1993) to the EMBL/GenBank/DBJ databases. H.				
REFERENCE		Deliuss, Deutsches Krebsforschungszentrum, Abteiling ATV, Im				
AUTHORS		Neuenheimer Feld 506, W 6900 Heidelberg, FRG				
TITLE		3 (bases 1 to 7855)				
JOURNAL		Deliuss, H. and Hofmann, B.				
MEDLINE		Primer-directed sequencing of human papillomavirus types				
FEATURES		Curr. Top. Microbiol. Immunol. 186, 13-31 (1994)				
source		Location/Qualifiers				
CDS		1..7855				
		/organism="Human papillomavirus type 26"				
		/clone="Insert in BamHI site of pUC19"				
		97...549				
		/gene="E6"				
		/codon_start=1				
		/product="early protein"				
		/db_xref="PID:g396957"				
		/translation="MEEDPRERTHELCESLNTLONLOVOCYCKETLOWADYVN				
		FAICDLRVYDRSPYACAKRCVIFSKITEYRYVCISYGTALBALTKNKSNLLIR				
		CHRCOMPLGPPEEKQRIVDKRRFRFETAGQWKGLCTNCWRPRROTQV"				
		556..870				
		/gene="E7"				
		/codon_start=1				
		/product="early protein"				
		/db_xref="PID:g396958"				
		/translation="MHGNIINIEDVLDLPQPEIDLCRYEQLDYEQFOSSDEDETDN				
		MRDQARQAGQEVYRIEACQCCNCSIVQLAVQSSRQNVRLFQMLMEDYSLVCHQCA				
		AO"				

[illegible]















```
Db 6072 gaggatagtaggacaaacattctgtgtgatccaaagcaaaacccaattgtgtattattggg 6131
QY 421 GAGGACGTTAGGACAAATGTCGTGTAGATTATAAGCAGACACAGATTATGTATTTTGGGC 480
Db 6132 tgcactctctataggaagaacattggcttaaaaggaacagctgctgtctgtgcgccct 6191
QY 481 TGTGCCCTGCTATTGGGGAACACTGGGCTAAAGGACACTGCTGTAAATCGCTCTTTA 540
Db 6192 gcaaaagggattctcctctggaactgttaattcctctatacagagatggcgcaatg 6251
QY 541 TCACAGGGCGATTGCCCCCTTTAGAACCTTAAAGAACACAGTTTGGAAAGTGTATAG 600
Db 6252 gttgatattgggtttgtgctcagatgatttaaaacattcaggaacttaaatcagatg 6311
QY 601 GTGATATCGATATGTCGCCATGATCTTAGTACATTCGAAGATCTAATGTGAGGTA 660
Db 6312 cctttagatattcaaatcccaacctgtaaatatcctgtattattataaaatgagtcagac 6371
QY 661 CCATGGGATATTGTGAGTCATATTGTGAATATCTCTGATATTACAAATGTCTGCAGAT 720
Db 6372 gctatggcgattctatgtggttctatttgcgtagggaaacagttatttgcctaggaactac 6431
QY 721 CTTATGGGATTCCTATGCTTTTGTCTTACGACGTGAGCAGCTTTTGTGTAGGCATTTT 780
Db 6432 tttaaatggcgaggtgctattgttgaaacaaattacacctagcacattatatataaaagtaca 6491
QY 781 TGGAAATAGGCGAGGTACTATGGGTGACACTGTGCTCAATCTTATATATATAAAGGCACA 840
Db 6492 aataacagggatccccgcgaagctcagtatattgtctactcctcctagtggtgctatggta 6551
QY 841 GGTATGCGTGCTTCACTGGCAGCTGTGTATTCTCCCTCTCCCAAGTGCTCTATTGTT 900
Db 6552 acctctagggctcagttgtttaaaacctactggttgcaacgcgcacagggacacaaat 6611
QY 901 ACCCTGACTCCCAGTTGTTTAATTAACCATATTTGTTTACATAAGGCACAGGGTCTATAC 960
Db 6612 aatggcattgttggggcaaacaggtattgttactgttggtagcacacactaggaacaca 6671
QY 961 AATGGTATCTGCTGGCATATCAATATTATTGTACTGTGGTAGATACCCTCTGTAGTACC 1020
Db 6672 aactgactactctgcaacacacaaagct-tat-cca- --ca-tataattcaaccaa 6725
QY 1021 AATTTACATATGTGCTCTACACAGCTCTCTGTACTCTGGGCAATATGATGTACCAAA 1080
Db 6725 attaaacagtatgtaagaacatgagaggaataatgaattcacagtttgttttaactgtgt 6785
QY 1081 TTAAAGCAGTATAGCAGACATGTTGAAGATATGATTTGCAGTTTATTTTTCAGTTATGT 1140
Db 6786 aaaaattcaactgtctgagaaactatggcctattttacatactactgaactcacttcaact 6845
QY 1141 ACTATTACTTTAACTGCAGATGTTATGTCCTATATTCAATAGTATGAATAGCAGTATTTA 1200
Db 6846 gggggctggaatatattggattgtcacccccagctgcccacacagcttagaggacaaatcaga 6905
QY 1201 GAGGATTGGAACTTTGGTGTCTCCCGCCGCCAACACTACTAGTTTGGTGGATACATATCGT 1260
Db 6906 tatgttaaagccttgcctatacctctgcaaaaggaacagcctcctgcctgtaaaaagggac 6965
QY 1261 TTTGTACAACTGTGTCTATTACCTGTCAAAAGGATGCTGCACCAGCTGAAAATAAGGAT 1320
Db 6966 ccaactagctaaataaagtttggatgttaactctacagacagcttttctgtctgacct 7025
QY 1321 CCCTATGATAAGTTAAAGTTTGGAAATGTGGATTTAAAGGAAAAGATTTCCTTTGGACTTA 1380
Db 7026 gaceaatcccaactggcgagaagtttttaattgcaacttggggttgcctactaaactct 7085
QY 1381 GATCAATATCCCTTGGACGTAAATTTTGTGTCAGGCTGGATTGCGTCGCAAGCCACC 1440
Db 7086 actactactaaaacgcctcgcccccagttctctctacctctacacactcagcacaag 7144
QY 1441 ATAGGCCCTCGTAAAGGTTCTGCTCCATCTGCCACTACGCTCTCTTAAACCTGCCAAGCG 1499
```

---

```
RESULT 13
LOCUS I13522 1517 bp DNA PAT 02-AUG-1995
DEFINITION Sequence 1 from patent US 5437951.
ACCESSION I13522
NID 996326
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1517)
AUTHORS Lowy,B.R., Schiller,J.T. and Kimbauer,R.
TITLE Self-assembling recombinant papillomavirus capsid proteins
JOURNAL Patent: US 5437951-A 1 01-AUG-1995;
FEATURES
source Location/Qualifiers
1..1517
BASE COUNT 487 a 293 c 284 g 453 t
ORIGIN
Query Match 34.4%; Score 524; DB 40; Length 1517;
Best Local Similarity 68.9%; Pred. No. 0.00e+00;
Matches 982; Conservative 0; Mismatches 440; Indels 3; Gaps 2;
Db 1 atgtctcttggctgcttagtgaggccactgtctacttgcctcctcctgtcccagtatctaaag 60
QY 1 ATGGGCTTTGTGGCGGCTAGTGACAATACCGTATACCTTCCACCTCCTCTCTGTGGCAAGA 60
Db 61 gtgtgaagcagggatgaatatgttcacgcacaaacatatattatcatcagcaggaacatcc 120
QY 61 GTTGTAAATACTGATGATTATGTGACTCGCACAAAGCATATTTTATCATGCTGGCAGCTCT 120
Db 121 agactactgcagtgaggaactccctatttctctattataaaacctaacaatacaaaaata 180
QY 121 AGATATTAACTGTGTGTAATCCATATTTTAGGGTTCCTCGAGGTGGTGGCAATAAGCAG 180
Db 181 tttagtcttaagatcaagattacaatacaggtatttagaatacatattaccttgcgccccc 240
QY 181 GATATTCCTTAGGTTCTGCATACCAATATAGATATTTTCGGGTGCAGTTACCTTGACCCA 240
Db 241 aataagttgtgtttctcgcacccctcatttataatccagatcacacagcggctggtttgg 300
QY 241 AATAAATTTGGTTTACCCTGATAATAGTATTTTATCTCTGAAACACACACGTTTAGTGTGG 300
Db 301 gacctgtaggtgttgaggtaggtcgtgctcagccattaggtgtggtggaattagtgccat 360
QY 301 GCCTGTGCTGAGTGGAAATGTGCCCTGTGTCAGCCTTTAGGTGTGGCCTTAGTGGGCAT 360
Db 361 cctttattaaataattggatgacacagaaaatgctagtgtctatgcagcaaatgcaggt 420
QY 361 CCATTTTAAATAAATTAGATGACACTGAAAGTTCCCATGCGCTACGCTCAATGTTTCT 420
Db 421 gggataatagagaatgtatatctatgattacaaaacaaacaaatgtgttttaattgggt 480
QY 421 GAGGACGTTAGGGACAATGTGTCTGTAGATTATTAAGCAGACACAGATTATGTTTGGGC 480
Db 481 tgcacaaccacctataggggaaacscctggggcaaaagatccccatgtaccaaatgttgcgta 540
QY 481 TGTGCCCTGCTATTGGGGAACACTGGGCTAAAGGCACACTGCTTGTAAATCGCGTCCCTTA 540
Db 541 aatccaggtgattgtccaccattagatttaataaacacagttattcaggaatgggtgatg 600
QY 541 TCACAGGGCGATTGCCCCCTTTAGAACTTAAGAACACAGTTTGTGGAAGATGGTGATAG 600
Db 601 gttcatactgcttgtgtgtatggccttactacattacaggtcaacaaagaatgaaagt 660
QY 601 GTAGATACTGGATATGTTGCCATGGACITTAGTACATTTGCAAGATACTAAATGTGAGGTA 660
Db 661 ccaactggatattgtacatctatttggaaaataatccagattattataaaatggtgtcagaa 720
QY 661 CCATTTGGATATTGTCAGTCTATTGTGTAATATCTCTGATTATTACAAATGTCCTGCAGAT 720
```





CC NCBI gi: 1098843 Location/Qualifiers  
 FH Key  
 FT source  
 FT 1..1484  
 FT /organism="Human papillomavirus"  
 FT /isolate="NM T446"  
 FT /note="collected by cervical swab at the University of New  
 FT Mexico, New Mexico, United States."  
 FT <1..>1484  
 FT /gene="L1"  
 FT /codon\_start=1  
 FT /product="late major capsid protein L1"  
 FT /db\_xref="PDB:1098844"  
 FT /translation="TVYLPVPVSVKVVSTDEYVARTNIYYHAGTSRLRAYGHPYFPIK  
 K PNNKILVPKVSGLQYRVERIHLPPDPNKFPGPDSFYNPDTQLRWACVGEVGRGQP  
 L GVIGSHPLINKLDDPENASAYANAGVDNREICSMYKQTLQCLGCKPPIGEHWCK  
 FT G SPTNVAVNPDCPPLELINTV IQDGDMDTGFAMDFTTLQANKSEVPLDICTSICK  
 FT Y PDYIKMVSPEYGSLSFPLRLRREQMVRHLFNRAAGVGENVPDDLVIKSGSGSTANLASS  
 FT N YFPTPSGSMYTSDAQIFNKYFLQRAQGHNGICWGNQLFVTVVDTTRSTRNMSLCRAI  
 FT S TSETTYKNTNFKEYLRHGEYDLQIFQLCKRITLADVMYIIHSMNSTILEDWNFGLQ  
 FT P PPGTLEDYTRFVTSQAIACQKHPPAPKEDPLKKYTFEVLNKEFSAADLDQFPLGR  
 FT K FLQAGLKAPKPTLGRKATPTTSSTSTAKRKRX"  
 FT S Q Sequence 1484 BP; 482 A; 285 C; 276 G; 441 T; 0 other;  
 Query Match 33.8%; Score 515; DB 15; Length 1484;  
 Best Local Similarity 69.0%; Pred. No. 0.00e+00;  
 Matches 964; Conservative 0; Mismatches 431; Indels 3; Gaps 2;  
 Db 1 actgtctactgctctgtccagtagtcttaaggtgttaagcagcgatgaatgtgtgca 60  
 Qy 28 ACCGTATACCTCCACCTCCCTCTGTGCAAGAGTGTGTAATACCTGATGATGTGACT 87  
 Db 61 cgcacacatattatcatcaggaacatccagactactgactgactgtgacatccctat 120  
 Qy 88 CGCAAGACATATTTATCATGCTGGCAGCTCTAGATTATTAACCTGTGGTAATCCATAT 147  
 Db 121 ttctctataaaacccataacataaataattagttctctaaagatcatcaggattacaa 180  
 Qy 148 TTAGGGTTCCTGCAGGTGGTGGCAATAGCAGGATATTCCTAGGTTCTGCATACCAA 207  
 Db 181 tacaggggtattagaatacattaccctgaccccaataagtttggtttccctgacacctca 240  
 Qy 208 TATAGATATTTTCGGGTGCTACCTACCTGACCCAAATAAATTTGGTTTACCTGATAATAGT 267  
 Db 241 ttctataatccagatacacagcggcgtgttggccctgtgaggtgtgaggtgaggtcgt 300  
 Qy 268 ATTTATAATCTGAAACACACACGTTAGTGTGGCCCTGTGCTGAGTGGAAATTTGGCCGT 327  
 Db 301 ggtcagccattaggtgtggcattagtgccatctttattataaataattggtgagaca 360  
 Qy 328 GGTACGCTTTAGGTGTGGCTTAGTGGGATCCATTTTATAATAATATAGATGACACT 387  
 Db 361 gaaaatgctagtctatcgacgaataatcagggtgtggataatagagaatgtatctatg 420  
 Qy 388 GAAAGTTCCCATGCGCTACGCTCAATGTTCTTGAGGACGTTAGGACAAATGTGCTCTA 447  
 Db 421 gattacaaaacacacattgtgttaatttggttgcaaacaccctattaggggaacactgg 480  
 Qy 448 GATTATAAGCAGACACAGTATATTTTGGCTGTGCCCCCTGTCTATTGGGGAACACTGG 507  
 Db 481 ggcaagatcccatgtaccaaatgtgcagtaaatccagtgattgtccaccattagag 540  
 Qy 508 GCTAAGGCACTGCTGTGTAATTCGGCTCTTTATCAGAGGCGGATGCCCCCTTTAGAA 567  
 Db 541 ttaataaacacagttattcaggatggtgatgtgattgactggcctttggtgctatggac 600

Qy 568 CTTAAGAACACAGTTTGGAGATGGTGATGTAGTACTGTGATATGGTCCATGGAC 627  
 Db 601 ttactacatcagcggcctaacaagaagttccactgactgatatattgtacattatgc 660  
 Qy 628 TTTAGTACATGCAAGATACTAATGTGAGTACCATGTGATATTGTGACTGATTGT 687  
 Db 661 aaatccagattatataaaatggtgtcagaaccatattgacagactattttttat 720  
 Qy 688 AATATCTGATTATTACAAATGCTGCAGATCTTATGGGATTCCATGTTTTTTCG 747  
 Db 721 ttacaaaggaacaaaatggtgttagacattatttaataagggtgtgtgtgtgtgaa 780  
 Qy 748 TTACGAGCTGAGCAGCTTTTGTAGCATTGTTGTAATAGGCGAGGTACTATGGTGAC 807  
 Db 781 aatgtacacacgatttatcattaaagctctggtgtctactgcaaaatttagcagttca 840  
 Qy 808 ACTGTGCTCAATCTTATATATTAAAGGCACAGGTATGCGTCTTCACCTGGCAGCTGT 867  
 Db 841 aattatttctcacactagtggttctatggtttacctctgtatgccccaaatattcaataa 900  
 Qy 868 GTGTATTCTCCCTCTCCAAGTGGCTCTATTGTTACCTCTGACTCCAGTGTGTTAATAA 927  
 Db 901 ccttattggtttacacagcagcagggccacaataatggtcatttgttgggttaacaaacta 960  
 Qy 928 CCATATTGGTTACATAAGGCACAGGGTCATAACAATGTTATCTGCTGCATATAATCAATTA 987  
 Db 961 ttcttactgtgtgtactacacgcagtcacaaatgtcattatgtgtgctacatctct 1020  
 Qy 988 TTTGTTACTGTGGTAGATACCACTCGTAGTACCAATTTAAACATATGTGCTTTCACACAG 1047  
 Db 1021 acttcagaaacta--ca--tataaaataactaaactttaagggatcctcacacatggggag 1077  
 Qy 1048 TCTCTGTACTGGGCAATATGATGCTACCAAAATTTAAGCAGTATAGCAGACATGTTGAA 1107  
 Db 1078 gaatatgatttacagtttatttttcaactgtgcataaacttaactgacagcagttatg 1137  
 Qy 1108 GAATGATGTTGCAAGTTATTTTTCAGTTATGTACTATTACTTTAACTGCGAGATGTTATG 1167  
 Db 1138 acatacatattctatgaattccactattttggagactgggaattttgtgtctacaaccc 1197  
 Qy 1168 TCCTATATTCTAGTATGAATAGCAGATTTTAGAGATTGGAACTTTGGTGTTCCTCCCC 1227  
 Db 1198 ccccgaggaggcacactagaaactataggtttgttaacatcccaggaatgtctgt 1257  
 Qy 1228 CCGCAACTACTACTGTTGGTGATACATATCGTTTGTACAACTGTGTCTATTACCTGT 1287  
 Db 1258 caaaacatacacctccagcacctaaagaagatcccccttaaaaaaacacactttttggaa 1317  
 Qy 1288 CAAAAGGATGCTGCACGAGTGAATAAGGATCCCTATGATAAGTTAAAGTTTGGAAAT 1347  
 Db 1318 gtaaatttaaggaagaaagttttctgcagacctagatcagtttcccttaggcagcgaattt 1377  
 Qy 1348 GTGATTAAAGGAAAAGTTTCTTTGGACTTAGTCAATATCCCCCTTGGAGCTAAATTT 1407  
 Db 1378 ttactacaagcaggattg 1395  
 Qy 1408 TTGTTTCAGGCTGGATTG 1425

Search completed: Thu Mar 27 14:02:19 1997  
 Job time : 1096 secs.





DNA sequences of the HPV39 Papilloma: virus genome - used as a probe for in-vitro diagnosis of Papilloma: virus infections in cervical, vulvar and penile cancer

Claim 1; Fig 1; 3pp; French.

DNA was isolated from a previously identified HPV39 clone (Virology 161, 374-384, 1987), subcloned and sequenced. The HPV39 genome was found to comprise 7833bp and have a G/C content of 40%. The genome comprises open reading frames conserved among all the known HPV's; the ORFs are postulated to code for the early proteins and capsid components of the virus. In addition, two ORFs have been found on the complementary strand of the genome (see features table). The non-coding regions contain three complete copies of an HPV enhancer as well as two degenerate versions of the consensus palindromic (location not given). Other putative features which are described in the specification as being present within the regulatory region but whose positions are not given include: 4 nuclear factor-1 binding sites, 2 activation protein-1 binding sites, a papillomavirus enhancer associated factor binding site and a glucocorticoid response element. The HPV39 sequence has most homology with HPV18. The HPV39 is thought to belong to a sub-group of genital HPV's (with HPV18, HPV45 and a new type cloned from carcinoma ME180 cell line) representing a potential oncogene.

Sequence 7833 BP; 2423 A; 1486 C; 1662 G; 2262 T;

Query Match 47.3%; Score 721; DB 4; Length 7833;  
Best Local Similarity 74.8%; Pred. No. 0.00e+00;  
Matches 1110; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

5643 atggctatgtggcgtgtgacagcagtggtgtattgctccacacctctgtggcgag 5702  
1 ATGGCTATGTGGCGGCTAGTGACAATACCGTATACCTCCACCTCTCTGTGGCAAGA 60

5703 gtgtgcaactgctgattatgtgtacacgcacagcagcatattattatgctggcagctct 5762  
61 GTTGTAAATACGTGATGATTATGTGCTCGCAAGCATATTTTATCATGCTGGCAGTCT 120

5763 agattattacacgaggacacacatttttaagctgggtatgaatgggtggtgcgacagcag 5822  
121 AGATTATTAACTGTGTAATCCATATTTAGGGTTCCTGCAGGTGGGCAATAAGCAG 180

5823 gacattcaaaagtgctcatatcaatagggtatttcggtgacattcccgatcct 5882  
181 GATATCTCTAGGTTCTGCATACCAATATAGAGTATTTTCGGTGCATGTACCTGACCCA 240

5883 aataaattcagattccagatgcactcttataataccagaacacacacgcttttagtagtg 5942  
241 AATAAATTGGTTTACCTGTAATAGTATTTATTAATCTCGAACAACACGTTTATGTGG 300

5943 gctgtgtaggggtgaggggtggtgggggggggggggggggggggggggggggggggggggg 6002  
301 GCGTGTCTGGAGTGGAAATGGCCGGTGGTGGGCTTTAGGTGTGGCCCTTAGTGGGCAAT 360

6003 ccattataataagacagggatgatacactgaatacactcaac---cattttcatcaaccacac 6059  
361 CCATTTTATAATAAATTAGATGACACTGAAAGTTCCATGCGCGTACGCTAATGTTTCT 420

6060 agggacagtagggataatgtgtgtgattataaaccagacacagttgtgattataggc 6119  
421 GAGGACGTTAGGGACAATGTGTCTAGATTATAGCAGACACAGTTATGATTATTGGGC 480

6120 tgtgtcccgccattggggagcactgggtgtaagggaagcagtcgaaccccaataagta 6179  
481 TGTGCCCTGCTATTGGGAACACTGGGCTAAAGGCACCTGCTGTAAATCGCTCCCTTA 540

6180 tctacgggggactgtccctcttgaaactagtaaacacccctattgaggtgggtgatg 6239  
541 TCACAGGCGGATGCCCCCTTTAGAACTTAAGAACACAGATTTTGGAGATGTTGATATG 600

6240 attgactgctgactgagctgactgttctgttcatttcaggaacacacacacaggggtg 6299  
601 GTAGATACTGGATATGGTCCATGGACCTTAGTACATTGCAAGATACATAATGTGAGGTA 660

6300 ccttttagatattgtcaatccattgtgaaatatactgattatttgcataatgtcgcagat 6359

661 CCATTGGATATTGTGTCAGTCTATTTGTTAAATATCCTGATTATTACAAATGCTCAGAT 720  
6360 gtctatggggagcagtagtct 6419  
721 CCTTATGGGGATCCCATGTTTTTTTTCCTACGACGCTGAGCAGCTTTTCTAGGCAATTT 780  
6420 tgaatacgtggtgtagtgggtgacgcacacctctctctctctctctctctctctctctctct 6479  
781 TGGATAGGCGCAGGTACTATGCGTGACACTGCGCTCAATCCCTATATATTAAGGCACA 840  
6480 gatatacgtgcaacccccggtagtctctctctctctctctctctctctctctctctctctct 6539  
841 GGATACGCTGCTCACTGCGACGCTGTGTGTTATCTCCCTCTCAAGTGGCTCTATTGTT 900  
6540 acctctgattcccgagttatttaataagcccttattggtcacataaagcccgagggccacaac 6599  
901 ACCTCTGACCTCCAGTTGTTTAAATAACCATATGTTGTTACATAAGGCACAGGGTCAATAAC 960  
6600 aatggtatgtgtggcacaatacatttcttctctctctctctctctctctctctctctctct 6659  
961 AATGATATCTGCTGGCATAATCAATTAATTTGTTACTGTGTAGATACCACTCGTAGTACC 1020  
6660 aactttacattatctacacctatagatgctctccatacactctctctctctctctctctct 6719  
1021 AATTTAAACAATATGCTTCTACACAGTCTCCTGTACCTGGCAATATGATGCTACCAAA 1080  
6720 tttaaggaataacacgagcagctggaggtgattgatttacaatttatttcaactgtgt 6779  
1081 TTTAAGCAGTATAGCAGACATGTTGAAGAATATGATTGTCAGTTTATTTTTCAGTTATGT 1140  
6780 actgtcacattaaacacatgattgttctctctctctctctctctctctctctctctctct 6839  
1141 ACTATTACTTTAACTGACAGATGTTATGCTCTATTTATCATAGTATGAATAGCAGTATTTTA 1200  
6840 gacaattggaattgtgtgtagctctccacacatctccagtttggtagacacattacaga 6899  
1201 GAGGATTGGAACATTGTTGTTCCCGCCGCCCACTACTAGTTTGGTGATACATATCGT 1260  
6900 taccatcagctgtgcagcattacatgtcaaaagactctccagcactccagcactcaaaagact 6959  
1261 TTTGTACATCTGTTGCTATCTGTCGAAAGGATGCTGCACACGCTGAAATAAAGGAT 1320  
6960 ccattatgcagctctaaagtttggaaagtgtgacttaagggaaagtttagttggaaactt 7019  
1321 CCTATGATAGTTAAAGTTTGGAAAGTGGATTAAAGGAAGATTTTCTTGGACTTA 1380  
7020 gatacattcccttgggacgttaaattttgttgacggccagggtccgcagggcgccctact 7079  
1381 GATCAATATCCCTTGGACGTAAATTTTGGTTCAGGCTGGATGCGTCGCAAGCCACC 1440  
7080 atagggtcccccgaagcgccctgctgcacacctctctctctctctctctctctctctctct 7123  
1441 ATAGGCCCTCGTAAACGTTCTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1484

RESULT 2  
ID Q47166 standard; DNA; 1517 BP.  
AC Q47166;  
DT 21-JAN-1994 (first entry)  
DE BPV1 L1 ORF.  
KW L1; capsid protein; bovine; human; papillomavirus; baculovirus;  
KW transfer vector; promoter; capsid protein; transformation; vaccine;  
KW neutralising antibody; vertebrate; ss.  
OS Bovine papillomavirus.  
PN US8032859-A.  
PD 15-JUL-1993.  
PF 03-SEP-1992; 941371.  
PR 03-SEP-1992; US-941371.  
PR 16-MAR-1993; US-032869.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
PI Kirmse R, Lowy DR, Schiller JT;  
DR WPI; 93-249995/31.



P-PSDB: R38807.  
 PT Recombinant papilloma virus capsid proteins - for vaccines  
 PS Example 1; Page 32-34; 45pp; English.  
 CC The sequences given in 047166-67 encode the L1 capsid proteins from  
 CC bovine and human papillomavirus respectively. These sequences may be  
 CC inserted into a baculovirus transfer vector and operatively  
 CC expressed by a promoter of the vector, and the capsid protein  
 CC produced by transformed cells. The capsid proteins may be used in  
 CC vaccines to induce high-titre neutralising antibody response in  
 CC vertebrates.  
 SQ Sequence 1517 BP; 487 A; 293 C; 284 G; 453 T;  
 Query Match 34.4%; Score 524; DB 7; Length 1517;  
 Best Local Similarity 68.9%; Pred. No. 0.00e+00;  
 Matches 982; Conservative 0; Mismatches 440; Indels 3; Gaps 2;

Db 1 atgtctcttggctgctagtagggccactgtctacttggctcctgtcccgatctatgaag 60  
 QY 1 ATGGCTTTGTGGGGCGCTAGTGAACAATACCGTATACCTTCCACCTCCTTCTGTGGCAAGA 60  
 Db 61 gtgtgaacacgcatgaataatgttgacgcacacaaacatatattatcatgcaggaacatcc 120  
 QY 61 GTTGTAAATGACTGATGATTAAGTACTGCGCACAAAGCATATTTATCATGTGGCAGCTCT 120  
 Db 121 agactactgtcagttggacatccctctatttccattataaaacacacacacacacacacata 180  
 QY 121 AGATTATTAACTGTGTGTAATCCATATTTTAGGGTTCCCTGCAGGTGGTGCATTAAGCAG 180  
 Db 181 ttggtcttaaaatcagcagattacacacaggggtatttagaatacattatcctgcacccc 240  
 QY 181 GATATTCTTAAGCTTTCTGCATACCAATATAGAGTATTTTCGGGTGAGCATTTACCTGACCCA 240  
 Db 241 aataagttgtgttctcagacacctctatttataatccagatcacacagcgctggtttgg 300  
 QY 241 AATAAATTTGGTTTACCTGATATAGTATTTAATCTTGAAACACAACTTTAGTGTGG 300  
 Db 301 gctgtgtaggtgttggtaggtcgtggtcagccattaggtgtggcattagtagtggccat 360  
 QY 301 GCCTGTGCTGGAGTGGAAATGCGCGTGGTGCAGCCTTTAGTGTGGCCTTAGTGGGCAT 360  
 Db 361 cctttattataaataatggatgacacagaaaatcgtagtgcttattcagcagaatgcaggt 420  
 QY 361 CCATTTTATAAATATAGTACACTGAAAGTTCCCATGCCGTACGTCTAATGTTCT 420  
 Db 421 gtgataatagaatgtatatctatgtgattacacacacacacacacacacacacacacacac 480  
 QY 421 GAGGACGTTAGGACAAATGCTGTGTAGATTAAGCAGACACAGATATGATTTTGGGC 480  
 Db 481 tgcaaacacacctataggggaacactggggcaagggatccccatgttaccacatgttgacgta 540  
 QY 481 TGTGCCCCCTGCTATTGGGGAACACTGGGCTAAGGACACTGCTTGTAAATCGCTCCTTTA 540  
 Db 541 aatccgggtgattgtccacattagattgattataaacaacagttattcaggatgtgtatg 600  
 QY 541 TCACAGGGCGGATGTCGCCCTTTTGAACACTTAAGAACACACAGTTTGGAAAGATGTGTATG 600  
 Db 601 gtccactagcgtttgtgtctggactttactacattacagcgttaacaaaagtgaagt 660  
 QY 601 GTAGACTGGATGTGTGCTGGACCTTTAGTACATTCGAAGATACTAAATGTGAGGTA 660  
 Db 661 ccactggatattgtacatctattgcaaatatccagatttatataaaatgggtgtcagaa 720  
 QY 661 CCATTGGATATTGTGACTGCTATTGTAAATATCCCTGATTATTACAAATGCTGCAGAT 720  
 Db 721 ccatatggcagacgttattttttatttaccaggaaggaacaaatgtttgttagacattta 780  
 QY 721 CCTATGGGGATTCATGTTTTTTTGTGTACGACGTGAGCAGCTTTTGTAGGCATTTT 780  
 Db 781 tttaatagggcgtgactgtgtgtaaatgtaccagacgatttatcatataaagcctct 840  
 QY 781 TGAATAGGCGAGGTACTATGTTGGTGACACTGTGCTCAATCCTTATATATAAAGGCACA 840

Db 841 gggctactgcaaaatttagccagtttcaaatattttctcacactagttggttctatggtt 900  
 QY 841 GGTATCGGTCTTACCTGGCAGCTGTGTATTTCTCCCTCTCCAAGTGGCTCTATTGTT 900  
 Db 901 acctctgagtcaccaaatattcaataaaaccttattgtttacaacagacagggccacaat 960  
 QY 901 ACCTCTGACTCCCAAGTGTGTTAATAAACCATATTTGTTTACATTAAGCAGAGGTCTATAAC 960  
 Db 961 aatggcattgttgggggttcaacacatttctactgttggtagtactcacacagcagta 1020  
 QY 961 AATGGTATCTGCTGGCATATCAATTTATTTGTACTGTGGTAGATACCACTCTGTATACC 1020  
 Db 1021 aatatgtcattatgtctgccatattacttactcagaacacta--ca-tataaaaataactaac 1077  
 QY 1021 AATTTAAACAATATGTCTTCTACACAGTCTCCTGTACCTGGGCAATATGATGCTACCAAA 1080  
 Db 1078 ttttaagagtagtacctacagacatggggagggaatgatgattacagtttttttcaactgtgc 1137  
 QY 1081 TTTAAGCAGTATAGCAGACATGTTGAAGAATATGATTTGCAGTTTATTTTTCAGTTATGT 1140  
 Db 1138 aaaaataaccttaactcagacgcttatgcacacacacacacacacacacacacacacac 1197  
 QY 1141 ACTATTACTTTAACTGACAGATGTTAAGTCTCTATATTCATAGTAGTAATAGCAGTATTTA 1200  
 Db 1198 gaggactggaattttgtctcacaacctccccccagggagggcacacacacacacacacacac 1257  
 QY 1201 GAGGATTGAACCTTGTGTGTTCCCCCCCCGCAACACTACTAGTTTGTGGATACATATCGT 1260  
 Db 1258 ttgttaacatccccaggaatgtgttgcataaaacacacacacacacacacacacacacacac 1317  
 QY 1261 TTTGTACAATCTGTGCTATTACCTGCTCAAAAGGATGCTGCACACAGCTGAAATAAGGAT 1320  
 Db 1318 ccccttcaaaaaacacacacttttggggaagtaaatataaaggaagttttcttcagacaccta 1377  
 QY 1321 CCTATGATAAGTTAAAGTTTGGATGTGGATTAAAGAAAGTTTCTTTTGGACTTA 1380  
 Db 1378 gatcagtttcttcttagcagcgaatattttactacaagcaggtattg 1422  
 QY 1381 GATCAATATCCCCCTGGAGCTAAATTTTGGTTTCAGGCTGGATTG 1425

## RESULT 3

ID Q47167 standard; DNA; 1517 BP.  
 AC Q47167;  
 DT 21-JAN-1994 (first entry)  
 DE HPV1 L1 ORF.  
 KW L1; capsid protein; bovine; human; papillomavirus; baculovirus;  
 KW transfer vector; promoter; capsid protein; transformation; vaccine;  
 KW neutralising antibody; vertebrate; ss.  
 OS Human papillomavirus.  
 PN US032869-A.  
 PD 15-JUL-1993.  
 PF 03-SEP-1992; 941371.  
 PR 03-SEP-1992; US-941371.  
 PR 16-MAR-1993; US-032869.  
 PA (USRH ) US DEPT HEALTH & HUMAN SERVICE.  
 PI Kirnbauer R, Lowy DR, Schiller JT;  
 DR WPI; 93-249995/31.  
 DR P-PSDB; R38808.  
 PT Recombinant papilloma virus capsid proteins - for vaccines  
 PT against papilloma virus and for diagnosis of virus infection  
 PS Example 1; Page 35-37; 45pp; English.  
 CC The sequences given in 047166-67 encode the L1 capsid proteins from  
 CC bovine and human papillomavirus respectively. These sequences may be  
 CC inserted into a baculovirus transfer vector and operatively  
 CC expressed by a promoter of the vector, and the capsid protein  
 CC produced by transformed cells. The capsid proteins may be used in  
 CC vaccines to induce high-titre neutralising antibody response in  
 CC vertebrates.  
 SQ Sequence 1517 BP; 487 A; 291 C; 286 G; 453 T;  
 Query Match 33.1%; Score 504; DB 7; Length 1517;  
 Best Local Similarity 68.2%; Pred. No. 0.00e+00;









[illegible]





```
QY 1026 AACATATGTCCTTACACAGCTCTCCTAGTCGGCAATATGATGCTACCAAAATTTAA 1085
Db 121 gcaactacagtagcagcagtgagggaatgattcacagttatatttcagttgagcactat 180
QY 1086 GCAGTATAGCAGACATGTTGAAGATATGATTTGCAGTTTATTTTTCAGTTATGACTAT 1145
Db 181 tactttaaactgagaggttatgtcctatccatgatacctacagtagtagtatattgga 240
QY 1146 TACTTTAACTGCAGATGTTATGCTCTATATCATAGTATGATGATGATGATGATGAT 1205
Db 241 ttggaatttggttgctacccacacacacacacacacacacacacacacacacacacac 300
QY 1206 TTGGAACCTTTGGTGTCCCGCCCGCCCACTACTAGTTGGTGGATACATATGCTTTGT 1265
Db 301 gcaatccgttgctgttacctgtcaaaaggatactacacacacacacacacacacacacac 360
QY 1266 ACATATCTGTTGCTATTTACCTGTCAAAAGGATGTCACCGCTGAATATGATGATGAT 1325
Db 361 tgataaattaaagtgttgactgttgacctaaggaagaaatttttctccgattt 414
QY 1326 TGATAAGTTAAAGTTTGGATGTGGATTTAAAGGAAAGTTTCTTTGGACTT 1379
```

RESULT 11 standard; DNA; 416 BP.

```
ID T10811
AC T10811;
DT 10-APR-1996 (first entry)
DE Human papilloma virus L1 region cervical carcinoma isolate C14.
KW Human papilloma virus; L1 region; isolate C14; amplified;
KW cervical carcinoma; ss.
OS Human papillomavirus.
PN US5447839-A.
PD 05-SEP-1995.
PF 09-SEP-1988; 243486.
PR 09-SEP-1988; US-243486.
PR 10-MAR-1989; US-322550.
PR 09-SEP-1989; WO-U03747.
PR 14-NOV-1990; US-613142.
PR 20-APR-1993; US-050743.
PA (HOFF ) HOFFMANN LA ROCHE INC.
PI Bauer HM, Greer CE, Manos MM, Resnick RM, Ting Y;
DR WPI; 95-319884/41.
DT Detection of human papilloma virus DNA by amplification - using
PT specific consensus primer pairs and pref. detection with generic or
PT type specific probes for use in research and diagnosis
PS Disclosure: Columns 13-14; 36pp; English.
CC The human papilloma virus (HPV) amplified L1 region isolates T10805
CC -T10812 were discovered using the HPV L1 region consensus primers
CC and probes of the invention. The isolate sequences can then be used
CC to generate primers and probes, for the specific detection of these
CC previously uncharacterised HPV types.
SQ Sequence 416 BP; 127 A; 74 C; 74 G; 141 T;
```

```
Query Match 17.5%; Score 266; DB 17; Length 416;
Best Local Similarity 82.7%; Pred. No. 1.07e-170;
Matches 343; Conservative 0; Mismatches 71; Indels 1; Gaps 1;

Db 1 tatttggcataaactcagttgttggtagttagtgagacactaccgcagtagtaattt 60
QY 966 TATCTGCTGGCATAAATCAATATTTTGTATGTTGGTAGATACCACTCGTAGTACCAATTT 1025
Db 61 aacattatggccctctacacaaaaactcgtgccaatacatatgatcactactaaattt 120
QY 1026 AACATATGTCCT-TCTACACAGTCTCCTGTACCTGGGCAATATGATGCTACCAAAATTTA 1084
Db 121 agcactatagtagacatgtggagggaatagattacagttatttttcagttgtagcacta 180
QY 1085 AGCAGTATAGCAGACATGTTGAAGATATGATTTGCGAGTTTATTTTTCAGTTATGACTA 1144
Db 181 ttactttaaactgagaggttatgccatataccatagtagtagtagtatatttggaa 240
QY 1145 TTACTTAACTGCAGATGTTATGCTCTATATCATAGTATGATGATGATGATGATGATGAT 1204
```

```
Db 241 attggaatttggtgtacctccacacactactcaaaagttagtgatatacatatcgttttg 300
QY 1205 ATTTGGAACCTTTGGTGTCCCGCCCGCCCACTACTAGTTTGGTGGATACATATCGTTTGG 1264
Db 301 tgcatacgttgctgttacctgtcaaaaggatactacacacacacacacacacacacacac 360
QY 1265 TACAATCTGTTGCTATTTACCTGTCAAAAGGATGTCGACCAGCTGAANAATTAAGGATCC 1324
Db 361 atgataaaattaaagtgttgactgttgacctaaggaagaaatttttctccgattt 415
QY 1325 ATGATAAGTTAAAGTTTGGATGTGGATTTAAAGGAAAGTTTCTTTGGACTT 1379
```

RESULT 12

```
ID T10805
AC T10805;
DT 10-APR-1996 (first entry)
DE Human papilloma virus L1 region isolate 36A a HPV59 variant.
KW Human papilloma virus; L1 region; isolate 36A; amplified;
KW HPV59 variant; ss.
OS Human papillomavirus.
PN US5447839-A.
PD 05-SEP-1995.
PF 09-SEP-1988; 243486.
PR 09-SEP-1988; US-243486.
PR 10-MAR-1989; US-322550.
PR 09-SEP-1989; WO-U03747.
PR 14-NOV-1990; US-613142.
PR 20-APR-1993; US-050743.
PA (HOFF ) HOFFMANN LA ROCHE INC.
PI Bauer HM, Greer CE, Manos MM, Resnick RM, Ting Y;
DR WPI; 95-319884/41.
DT Detection of human papilloma virus DNA by amplification - using
PT specific consensus primer pairs and pref. detection with generic or
PT type specific probes for use in research and diagnosis
PS Disclosure: Columns 11-12; 36pp; English.
CC The human papilloma virus (HPV) amplified L1 region isolates T10805
CC -T10812 were discovered using the HPV L1 region consensus primers
CC and probes of the invention. The isolate sequences can then be used
CC to generate primers and probes, for the specific detection of these
CC previously uncharacterised HPV types.
SQ Sequence 455 BP; 131 A; 88 C; 79 G; 144 T;
```

```
Query Match 16.5%; Score 251; DB 17; Length 455;
Best Local Similarity 74.9%; Pred. No. 1.41e-159;
Matches 341; Conservative 12; Mismatches 102; Indels 0; Gaps 0;

Db 1 gmcagggcataaagtgttatgttgagacacacacacacacacacacacacacacacacac 60
QY 946 GCHAGGTCATACAAAGTATCTGCTGGCATATCATATTTTGTACTGTGTAGAT 1005
Db 61 actactcgcagcacacaactctgtgtgtgtcttactactcttcttcttacttacttactgaa 120
QY 1006 ACCACTCGTAGTACCAATTTTAAATATATGCTCTTACACAGTCTCCTGTACTGGGCAA 1065
Db 121 tacacacctaccagttttaagaataatgccagacatgtgrrggaatttgattgcagttt 180
QY 1066 TATGATGTCTACCAATTTAAAGCAGTATATACAGACATGTTGAAGAATATGATTTGCAGTT 1125
Db 181 atayttcaactgtgttaaaataacacwtaactacagagtaagtgcatacatcataaata 240
QY 1126 ATTTTTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1185
Db 241 aatcaccattttggaggaggttggaattttgtttttacacacacttacttactgactartta 300
QY 1186 AATAGCAGTATTTTAGAGGATTGGAACCTTTGTGTTCCTCCCGCCGCAACTACTAGTTG 1245
Db 301 gttgcacacatccgtttttgttcaatttctgctgtgtaacttgcataaagacacccgcag 360
QY 1246 GTGGATACATATCGTTTGTGACAAATCTGTGTGCTATTACCTGTCTAAAGGATGCTGCACCA 1305
Db 361 ccagtttaaacaggagccctttatgacaaactaaagtgttgagactgtaaaacttaaggaa 420
```



QY 1306 GCTGAAAATAGGATCCCTATGATAAGTTTGAATGGTGGATTTAAAGGAAAAG 1365  
 Db 421 ttttgcagatcttgatcagtwcccytkggacg 455  
 QY 1366 TTTTCTTTGGACTTAGATCAATATCCCTGGACG 1400

## RESULT 13

ID Q56374 standard; DNA; 455 BP.  
 AC Q56374;  
 DT 29-JUL-1994 (first entry)  
 DE HPV isolate 36A LI amplified region.  
 KW Human papilloma virus; amplification; polymerase chain reaction;  
 KW PCR; detection; assay; ss.  
 OS Human papilloma virus isolate 36A.  
 PN US283171-A.  
 PD 01-FEB-1994.  
 PF 29-AUG-1989; 651356.  
 PR 09-SEP-1988; US-243486.  
 PR 10-MAR-1989; US-322550.  
 PR 29-AUG-1989; WO-U03747.  
 PR 15-FEB-1991; US-651356.  
 PA (HOFF) HOFFMANN LA ROCHE INC.  
 PA (UIRP) UNIV ROCHESTER.  
 PI Broker TR, Manos MM, Ting Y, Wolinsky SM, Wright DK.  
 DR WPI; 94-048082/06.  
 DT Detection of genital human papilloma virus - by PCR amplification  
 PT using defined consensus primer pairs  
 PS Disclosure; Page 7; 13pp; English.  
 CC The sequence is that of the LI amplified region of HPV isolate 36A.  
 SQ Sequence 455 BP; 131 A; 88 C; 79 G; 144 T;

Query Match 16.5%; Score 251; DB 10; Length 455;

Best Local Similarity 74.9%; Pred. No. 1.41e-159; Indels 0; Gaps 0;  
 Matches 341; Conservative 12; Mismatches 102;

Db 1 gmcaggwgcataaayaatggtatattgttgccacaaatcaattgttttaacagtttagat 60  
 QY 946 GCACAGGGTCATAACAAATGCTATCTGCTGCATAAATTAATTTGTTACTGTGGTAGAT 1005  
 Db 61 actactgcgcagcccaatctyctgtgtgtgtcttactactctctctctcctaatgaa 120  
 QY 1006 ACCACTCGTAGTACCAATTTAAACAATATGCTTCTACACAGTCTCTGTACCTGGGCAA 1065  
 Db 121 tacacaccaccagttttaaagaatgccagacatgtggrggaatttgatggcagttt 180  
 QY 1066 TATGATGCTACCAATTTAAAGCAGTAGACAGACATGTTGAAGAATATGATTGCGAGTTT 1125  
 Db 181 atayttcaactgtgttaaataacwttaactacagaggttaattgtcacaatcataatg 240  
 QY 1126 ATTTTCAGTTATGTACTATTACTTAACTGCAGATGTTTAAAGAAATATGATTGCGAGTTT 1185  
 Db 241 aataccactatttggagattggaattttgtrttacaccactctctactctacttata 300  
 QY 1186 AATAGCAGTATTTTAGAGGATTTGGAATTTGGTGGTTCCTCCCGCCGCAACTACTAGTATG 1185  
 Db 241 aataccactatttggagattggaattttgtrttacaccactctctactctacttata 300  
 QY 1186 AATAGCAGTATTTTAGAGGATTTGGAATTTGGTGGTTCCTCCCGCCGCAACTACTAGTATG 1245  
 Db 301 gttgacacataccgttttgcatactgtgtgtaactgtgtcaaaaggacaccgcacg 360  
 QY 1246 GTGGATACATATCGTTTGTGACAAATGTTGCTATTACCTGTCTCAAAAGGATGCTGCACCA 1305  
 Db 361 ccagtttaaacagagacccttatgacaaactaaagttttggactgttaactcttaaggaaag 420  
 QY 1306 GCTGAAAATAGGATCCCTATGATAAGTTTGAATGGTGGATTTAAAGGAAAAG 1365  
 Db 421 ttttgcagatcttgatcagtwcccytkggacg 455  
 QY 1366 TTTTCTTTGGACTTAGATCAATATCCCTGGACG 1400

## RESULT 14

ID Q03903 standard; DNA; 455 BP.  
 AC Q03903;

DT 24-AUG-1990 (first entry)  
 DE LI amplified region; isolate 36A.  
 KW Papilloma-virus; consensus primer; PCR; probe; ss.  
 OS Synthetic.  
 PN WO9002821-A.  
 PD 22-MAR-1990.  
 PF 29-AUG-1989; 03747.  
 PR 10-MAR-1989; US-322550.  
 PR 10-MAR-1989; US-243486.  
 PA (CETU) Cetus Corp.  
 PI Manos MM, Wright DK, Ting Y, Broker TR.  
 DR WPI; 90-116005/15.  
 DT Detecting and priming human papilloma-virus - using consensus primers  
 PT in polymerase chain reaction to amplify particular genomic regions.  
 PS Disclosure; Page 10; 33pp; English.  
 CC The sequence is markedly different than the published sequence for HPV's.  
 CC Discovered using LI consensus probe hybridising to amplified DNA.  
 CC The sequence enables one to isolate the entire virus from samples  
 CC containing the virus.  
 CC See also Q03898-Q03949.  
 SQ Sequence 455 BP; 131 A; 88 C; 79 G; 144 T; 13 Others;

Query Match 16.5%; Score 251; DB 1; Length 455;

Best Local Similarity 74.9%; Pred. No. 1.41e-159; Indels 0; Gaps 0;  
 Matches 341; Conservative 12; Mismatches 102;

Db 1 gmcaggwgcataaayaatggtatattgttgccacaaatcaattgttttaacagtttagat 60  
 QY 946 GCACAGGGTCATAACAAATGCTATCTGCTGCATAAATTAATTTGTTACTGTGGTAGAT 1005  
 Db 61 actactgcgcagcccaatctyctgtgtgtgtcttactactctctctcctaatgaa 120  
 QY 1006 ACCACTCGTAGTACCAATTTAAACAATATGCTTCTACACAGTCTCTGTACCTGGGCAA 1065  
 Db 121 tacacaccaccagttttaaagaatgccagacatgtggrggaatttgatggcagttt 180  
 QY 1066 TATGATGCTACCAATTTAAAGCAGTAGACAGACATGTTGAAGAATATGATTGCGAGTTT 1125  
 Db 181 atayttcaactgtgttaaataacwttaactacagaggttaattgtcacaatcataatg 240  
 QY 1126 ATTTTCAGTTATGTACTATTACTTAACTGCAGATGTTTAAAGAAATATGATTGCGAGTTT 1185  
 Db 241 aataccactatttggagattggaattttgtrttacaccactctctactctacttata 300  
 QY 1186 AATAGCAGTATTTTAGAGGATTTGGAATTTGGTGGTTCCTCCCGCCGCAACTACTAGTATG 1245  
 Db 301 gttgacacataccgttttgcatactgtgtgtaactgtgtcaaaaggacaccgcacg 360  
 QY 1246 GTGGATACATATCGTTTGTGACAAATGTTGCTATTACCTGTCTCAAAAGGATGCTGCACCA 1305  
 Db 361 ccagtttaaacagagacccttatgacaaactaaagttttggactgttaactcttaaggaaag 420  
 QY 1306 GCTGAAAATAGGATCCCTATGATAAGTTTGAATGGTGGATTTAAAGGAAAAG 1365  
 Db 421 ttttgcagatcttgatcagtwcccytkggacg 455  
 QY 1366 TTTTCTTTGGACTTAGATCAATATCCCTGGACG 1400

## RESULT 15

ID T03499 standard; DNA; 647 BP.  
 AC T03499;  
 DT 12-JUN-1996 (first entry)  
 DE Papilloma virus major capsid protein DNA  
 KW HP-virus 29; papilloma virus; major;  
 KW capsid protein; plasmid VS93-1; DSM 9133; diagnosis;  
 KW skin carcinomas; therapy; vaccination; ds.  
 OS Papilloma virus.  
 FH Key Location/Qualifiers  
 FT mat\_peptide 1..645  
 FT /\*tag= a  
 PN DE4415743-A1.  
 PD 09-NOV-1995.

```
PF 04-MAY-1994; 415743.
PR 04-MAY-1994; DE-415743.
PI (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI: 95-383680/50.
DR P-PSDB: R88267.
PT DNA encoding peptide(s) of papilloma virus major capsid protein -
PT useful for detecting papilloma virus in skin carcinoma
PS Claim 1: Fig 1: 15pp, German
CC T03499 is 82.7% homologous to HP-virus 29, encodes R88267 a
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS93-1 (DSM 9133). The DNA is useful in
CC diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
CC Sequence 647 BP; 167 A; 157 C; 157 G; 166 T;
SQ
```

Query Match 13.7%; Score 209; DB 18; Length 647;  
Best Local Similarity 66.3%; Pred. No. 1.42e-128;  
Matches 425; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Db 7 ggacagccattaggcgtgggttaagtggacacccctctgtatacaaaactgaatgacact 66  
QY 328 GGTCAAGCCTTTAGTGTGGCCTTAGTGGGATCCATTTATAATAAATTAGATGACACT 387  
Db 67 gaaaactccaacattgcacatgctgacaatagctcctgactcccgaggacacaaacttctgtt 126  
QY 388 GAAAGTTCCCATGCGGCTACGCTAATGTTCTGAGGACGTTAGGGACAAATGTCGTGTA 447  
Db 127 gactgaagcaacacacactgtgcatactactggctgtacgcccccatgggggaatactgg 186  
QY 448 GATTATAAGCAGACACAGTATGTTATTTGGGCTGTGCCCTGCTATTGGGGAACACTGG 507  
Db 187 ggtaaagggtaccctctgtgcagctactaactacccacaggagactgctccctctggag 246  
QY 508 GCTAAGGCACTGCTGTAAATCGCGTCTTTATACAGGGCGATTGCCCCCTTTAGAA 567  
Db 247 ttaatgacatctattatcagatagcgacatggtgataccggatggtgccatggac 306  
QY 568 CTTAAGACACAGTTTGGAGATGGTATGTTAGTACTGGATATGTTGCCATGGAC 627  
Db 307 tttaactgccctgaatttaattgaattgaactgacgtgcccctgataattggcagctatttgc 366  
QY 628 TTTAGTACTATGCAAGTACTAAATGTGAGTACCATTTGGATTTGTCATGTCATTTGT 587  
Db 367 aaatatcccgattatttgggcatggtgcgaccccgatgacgatagcatagttctcttttc 426  
QY 688 AAATATCCTGATTATTACAAATGTCAGATCCTTATGGGATTCATGCTTTTTCG 747  
Db 427 ctccgtcgggaacaactgtttgcccagacacttttcaatcgt-gcgggtgagtgttgagac 486  
QY 748 TTACGACGTGAGCAGCTTTTGTAGGCATTTTGGAAATAGGCGCAGGTACTATGGGTGAC 807  
Db 487 aaaaattccagaatctttgtacacctcaaaagggagtagcgggcggtgagactcccggcagtgct 546  
QY 808 ACTGTGCCTCAATCCTTATATATTAAGGACAGGTATGCGTCTTCACCTGGCAGCTGT 867  
Db 547 atatacagccccacacccagtggtgtctatggtgacctctgagggcacaaaatttcaataag 606  
QY 868 GGTATTCTCCCTCTCCAGTGGCTATTTGTTACCTCTGACTCCCAAGTTGTTTAAATAA 927  
Db 607 tcttaactggctcagcagctcaagcccaaaataacggtat 647  
QY 928 CCATATGTTTACATAAGGCACAGGGGTATAACAATGCTAT 968

Search completed: Thu Mar 27 14:05:21 1997  
Job time : 163 secs.

\*\*\*\*\*

W P S R E L F

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPSrch\_ntp n.a. - n.a. Smith-Waterman search, using a protein database  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Mar 26 08:44:39 1997; MasPar time 111.14 seconds  
Tabular output not generated. 847.073 Million cell updates/sec

Title: >US-08-409-122-1  
Description: (1-1524) from US08409122.seq  
Perfect Score: 7620  
N.A. Sequence: 1 ATGGCTTTGTGGCGGCTAG.....GTGTAGTGGCCAGAGTAA 1524  
Comp: TACGAAACCCCGGATC.....CACATGACGGTCTTCATT

Scoring table: TABLE bktranslated2  
Gap 30

Nmatch STD : Dbase 0; Query 0  
Searched: 88003 seqs, 3088698 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq25

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18

Statistics: Mean 58.520; Variance 269.547; scale 0.217

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	3582	47.0	505	8 R38807	BPV1 L1.	0.00e+00
2	3456	45.4	505	8 R38808	HPV1 L1.	4.89e-236
3	3049	40.0	532	17 R83016	Recombinant papilloma	2.69e-258
4	1972	25.9	495	3 P51099	Sequence of bovine pa	5.81e-159
5	1809	23.7	508	4 P30315	Sequence encoded by t	4.96e-144
6	1739	22.8	215	16 R88267	Papilloma virus major	1.25e-137
7	821	10.8	221	16 R88273	Papilloma virus major	7.11e-55
8	808	10.6	228	16 R88275	Papilloma virus major	9.93e-54
9	797	10.5	220	16 R88269	Papilloma virus major	9.22e-53
10	780	10.2	224	16 R88271	Papilloma virus major	2.88e-51
11	762	10.0	220	16 R88272	Papilloma virus major	1.09e-49
12	756	9.9	222	16 R88268	Papilloma virus major	3.66e-49
13	706	9.3	225	16 R88270	Papilloma virus major	8.57e-45
14	684	9.0	224	16 R88274	Papilloma virus major	7.04e-43
15	653	8.6	88	6 R82057	Rabbit serum reactive	3.44e-40
16	653	8.6	88	3 R14299	Seroreactive epitope	3.44e-40
17	448	5.9	39	1 P82074	Human papilloma virus	9.28e-23

18	381	5.0	36	6 R32071	VLP reactive HPV16 L1	2.96e-17
19	347	4.6	36	6 R32065	VLP reactive HPV16 L1	1.62e-14
20	345	4.5	35	6 R32063	Human IGA reactive HP	2.34e-14
21	328	4.3	45	6 R32070	VLP reactive HPV16 L1	5.23e-13
22	313	4.1	51	6 R32062	Human IGA reactive HP	7.93e-12
23	313	4.1	51	6 R32059	Human IGA reactive HP	7.93e-12
24	262	3.4	21	3 R15589	Immunopeptide derived	6.62e-08
25	248	3.3	36	6 R32066	VLP reactive HPV16 L1	7.37e-07
26	231	3.0	20	1 R04980	Papilloma virus type	1.31e-05
27	227	3.0	20	1 R04984	Papilloma virus type	2.56e-05
28	223	2.9	36	6 R32069	VLP reactive HPV16 L1	4.98e-05
29	212	2.8	20	1 R04981	Papilloma virus type	3.05e-04
30	212	2.8	20	6 R32058	Human IGA reactive HP	3.05e-04
31	212	2.8	20	1 R04967	Papilloma virus type	3.05e-04
32	212	2.8	20	6 R32061	Human IGA reactive HP	3.05e-04
33	211	2.8	20	1 R04955	Papilloma virus type	3.59e-04
34	210	2.8	20	1 R04985	Papilloma virus type	4.23e-04
35	217	2.8	26	6 R32068	VLP reactive HPV16 L1	1.34e-04
36	205	2.7	20	1 R04966	Papilloma virus type	9.52e-04
37	205	2.7	20	1 R04974	Papilloma virus type	9.52e-04
38	201	2.6	21	6 R32060	Human IGA reactive HP	1.81e-03
39	197	2.6	37	3 R14301	Seroreactive epitope	3.44e-03
40	191	2.5	20	1 R04973	Papilloma virus type	8.90e-03
41	191	2.5	20	1 R04964	Papilloma virus type	8.90e-03
42	193	2.5	39	1 P82075	Hman papilloma virus	6.49e-03
43	185	2.4	20	1 R04976	Papilloma virus type	2.28e-02
44	185	2.4	20	1 R04990	Papilloma virus type	2.28e-02
45	184	2.4	88	3 R14298	Seroreactive epitope	2.67e-02

ALIGNMENTS

RESULT 1  
ID R38807 standard; Protein; 505 AA.

AC R38807;  
DE 21-JAN-1994 (first entry)  
DE BPV1 L1.  
KW L1; capsid protein; bovine; human; papillomavirus; baculovirus;  
KW transfer vector; promoter; capsid protein; transformation; vaccine;  
KW neutralising antibody; vertebrate.  
OS Bovine papillomavirus.  
PN US8032869-A.  
PD 15-JUL-1993.  
PF 03-SEP-1992; 941371.  
PR 03-SEP-1992; US-941371.  
PR 16-MAR-1993; US-032865.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
PI Kirnbauer R, Lowy DR, Schiller JT;  
DR WPI; 93-249995/31.  
DR N-PSDB; Q47166.  
PT Recombinant papilloma virus capsid proteins - for vaccines  
PT against papilloma virus and for diagnosis of virus infection  
PS Example 1; Page 32-34; 45pp; English.  
CC The sequences given in R38807-08 represent the L1 capsid proteins from  
CC bovine and human papillomavirus respectively. The DNA encoding these  
CC sequences may be inserted into a baculovirus transfer vector and  
CC operatively expressed by a promoter of the vector, and the capsid  
CC protein produced by transformed cells. These capsid proteins may be  
CC used in vaccines to induce high-titre neutralising antibody response  
CC in vertebrates.  
SQ Sequence 505 AA;

Query Match 47.0%; Score 3582; DB 8; Length 505;  
Best Local Similarity 47.7%; Pred. No. 0.00e+00;  
Matches 709; Conservative 285; Mismatches 489; Indels 3; Gaps 2;

Db m s l w l p s e a t v i l p p v p v s k  
Dt 1 atgwnyntnggycnwgargcnagctgtataytncnccngtncngtwnaar 60  
Qy 1 ATGGCTTTGTGGCGGCTAGTGCACAAATACCTTACCTCTTCTGGCAAGA 60  
Qt M A L W R P S D N T V Y L P P P S V A R  
Db v v s t d e y v a r t n i y y h a g t s

















```

QY 502 CACCTGCTAAAGCGCATCTGTTAAATCCGCTCTTATCACA-GGGCGAATGCCCC 560
Qt H W A K G T A C K S V L Y H X G D C P P

Db 240 nyngarynaaraaywngnathargayngngaygatgtyayathngnttygnaa 299
Dt 1 e l k n s v i e d g d m f d i g f g n

QY 561 TTTAGAACCTAAGACACACAGTTTGAAGATGTGTATGGTAGACAGTGGATGTGC 620
Qt L E L K N T V L E D G D M V D T G Y G A

Db 300 i n d n k t l s f n r s d v s l d i v n e
Dt 3 a t h a a y a a r a c n y t w n t t a y n g n w n g n y t n g a y a t h n g n a y g a 359
QY 621 CATGGAGCTATCTCATCAATTAATAGGATGAGTACCATTCGATATTTGTGTCAG 680
Qt M D F S T L Q Q D T K C E V P L D I C Q S

Db 360 r a c h t y a a r t a y c n g a y t t y t n a c n a t g w n a a y g a y g t n a y g n g a y w n t g y t t 419
Dt 1 c k y p d f l t m s n d v y g d s c f
QY 681 TATTTGTAATATCCGTGATTAATTAATCAATGTCTGCAGATCCCTATGGGANTCCATGT 740
Qt I C K Y P D Y L Q M S A D P Y G D S M F

Db 420 y t t y t g n g n g n g a r c a r t g y t a y c n g n c a y t a y t t y g n g n g n g n g n t g t 479
Dt f c a x r e q c y a x h y f v x g g v v
QY 741 TTTTCTTACGACGACGAGCTTTTCTAGCATTTTGGATTTTGGATAGGCGAGTACTAT 800
Qt F C L R R E Q L F A R H F W N R A G T M

Db 480 n g n g a y w n a t h c c 494
Dt 1
QY 801 GGGTACACTGTGCC 815
Qt G D T V

RESULT 8
ID R88275 standard; Peptide; 228 AA.
AC R88275;
DT 12-JUN-1996 (first entry)
DE Papilloma virus major capsid protein.
KW HP-Virus 12; papilloma virus; major;
KW capsid protein; plasmid VS75-3; DSM 9141; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PR (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI: 95-383680/50.
DR N-PSDB; T03507.
PT DNA encoding peptide(s) of papilloma virus major capsid protein
PT useful for detecting papilloma virus in skin carcinoma
PS Claim 7; Fig 9; 15pp; German.
CC T03507 is 75.7% homologous to HP-Virus 12, encodes R88275 a
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS75-3 (DSM 9141). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid
CC proteins encoded by a PV genome contg. the DNA are useful for
CC therapy and/or vaccination.
SQ Sequence 228 AA;

```

```

Query Watch 10.68; Score 808; DB 16; Length 228;
Best Local Similarity 41.94; Pred. No. 9.93e-54;
Matches 205; Conservative 96; Mismatches 185; Indels 3; Gaps 3;

Db 1 g n m g n g n c a r c n y t n g n g n g n a c n w n g n c a y c n y t n t t a y a a r g n a y 60
Dt 1 e l k n s v i e d g d m f d i g f g n

QY 322 GGGCGTGGCAGCTTAGTGTGGCTTAGTGGCGATCCATTTATTAATATAGAT 381
Qt G R G Q P L G V G L S G H P F Y N K L D

```

```

Db 61 g a y g n c a r a a y c n y t n g n c a r g n c n t t y w s n a c n g a y g a y m n c a r a y 120
Dt 1 e l k n s v i e d g d m f d i g f g n
QY 382 GACACTGAAGTCCCATCCGCTAGCTAATG-TTTCG-A-GACGTTAGGACAT 438
Qt D T E S S H A A T S N X F X X D V R D N

Db 121 a c n w n t t y g a y c n a r a t h c a r t g t t a t h g n t g y g n c n c n t g a t h g n 180
Dt t s f d p k q i q m f i i g c a p c i g
QY 439 GGTCTGTAGATATAAGCAGACAGTATATTTGGCTGTCCCTGCTATTGG 498
Qt V S V D Y K Q T Q L C I L G C A P A I G

Db 181 g a r c a y t g g a y t n g n g a r m n g t g c n g n g n a a y a a y a a y g n g n m n t g y c n 240
Dt e h w d v g e r c a g a n n e n g r c p
QY 499 GAACACTGGCTAAAGGCACTGCTGTAAATCGCTTATCAGAGGGCGATTCGCC 558
Qt E H W A K G T A C K S R P L S Q G D C P

Db 241 c n a t h a a r y t n a a y w n g n a t h c a r g a y g n g a y a t g n g a y a t h g n t a y g n 300
Dt p i k l v n s v i q d g d m a d i g y g
QY 559 CCTTTAGAACTTAAGACACAGCTTTTGGAGATGGTGATATGTAGTACTGATATG 618
Qt P L E L K N T V L E D G D M V D T G Y G

Db 301 a a y t n a a y t n g n a c n y t n c a r a a y m g n w n g a y t n w n y t n g a y a t h g n a y 360
Dt 1 e l k n s v i e d g d m f d i g f g n
QY 619 GCATGGACTTAGTACATTCGACATTAATGAGGAGGTACCATTCGATATTTCTAG 678
Qt A M D F S T L Q D T K C E V P L D I C Q

Db 361 g a r a c n t g y a a r t a y c n g a y t t y t n a a r a t g c a r a a y a t h t a y g n g a y w n t g y 420
Dt e t c k y p d f l k m q n d i y g d s c
QY 679 TCTATTTGTAATATCCTGATTAATTTACAAATGCTGACAGATCCTTATGGGATTCATG 738
Qt S I C K Y P D Y L Q M S A D P Y G D S M

Db 421 t t y t t y t g n g n g n g a r c a r t g y t a y c n g n c a y t t y t n g n g n g n g n a a r 480
Dt 1 e l k n s v i e d g d m f d i g f g n
QY 739 TTTTGTGTTAGCAGCTGACGCTTTTGTAGCATTTTGGATTTGGAATAGGCGACTACT 798
Qt F F C L R R E Q L F A R H F W N R A G T

Db 481 g n g n g a y 489
Dt a g d
QY 799 ATGGGTGAC 807
Qt M G D

RESULT 9
ID R88269 standard; Peptide; 220 AA.
AC R88269;
DT 12-JUN-1996 (first entry)
DE Papilloma virus major capsid protein.
KW HP-Virus 49; papilloma virus; major;
KW capsid protein; plasmid VS40-7; DSM 9135; diagnosis;
KW skin carcinomas; therapy; vaccination.
OS Papilloma virus.
PN DE4415743-A1.
PD 09-NOV-1995.
PF 04-MAY-1994; 415743.
PR (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI de Villiers zur Hausen E, Leigh I, Shamanin V, zur Hausen H;
DR WPI: 95-383680/50.
DR N-PSDB; T03501.
PT DNA encoding peptide(s) of papilloma virus major capsid protein
PT useful for detecting papilloma virus in skin carcinoma
PS Claim 7; Fig 3; 15pp; German.
CC T03501 is 78.5% homologous to HP-Virus 49, encodes R88269 a
CC peptide of papilloma virus (PV) major capsid protein and is
CC expressed by the plasmid VS40-7 (DSM 9135). The DNA is useful
CC in diagnosis, esp. to detect PV in skin carcinomas. Major capsid

```









**THIS PAGE BLANK (USPTO)**